The Economics of Non-Market Goods and Resources

Patricia A. Champ
Kevin J. Boyle
Thomas C. Brown *Editors*

A Primer on Nonmarket Valuation

Second Edition

Springer

The Economics of Non-Market Goods and Resources

Volume 13

Series editor

I.J. Bateman, School of Environmental Sciences, University of East Anglia, Norwich, UK

More information about this series at http://www.springer.com/series/5919

Patricia A. Champ · Kevin J. Boyle
Thomas C. Brown
Editors

A Primer on Nonmarket Valuation

Second Edition

Springer

Editors
Patricia A. Champ
U.S. Forest Service, Rocky Mountain Research Station
Fort Collins, CO
USA

Thomas C. Brown
U.S. Forest Service, Rocky Mountain Research Station
Fort Collins, CO
USA

Kevin J. Boyle
Virginia Tech
Blacksburg, VA
USA

ISSN 1571-487X
The Economics of Non-Market Goods and Resources
ISBN 978-94-007-7103-1   ISBN 978-94-007-7104-8 (eBook)
DOI 10.1007/978-94-007-7104-8

Library of Congress Control Number: 2016958973

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media B.V.
The registered company address is: Van Godewijckstraat 30, 3311 GX Dordrecht, The Netherlands

Contents

Contributors

Wiktor L. Adamowicz University of Alberta, Edmonton, AB, Canada

Richard C. Bishop University of Wisconsin-Madison, Madison, WI, USA

Kevin J. Boyle Virginia Tech, Blacksburg, VA, USA

Thomas C. Brown U.S. Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA

Fredrik Carlsson University of Gothenburg, Gothenburg, Sweden

Patricia A. Champ U.S. Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA

Mark Dickie University of Central Florida, Orlando, FL, USA

Nicholas E. Flores University of Colorado-Boulder, Boulder, CO, USA

Thomas P. Holmes U.S. Forest Service, Southern Research Station, Research Triangle Park, NC, USA

Craig E. Landry University of Georgia, Athens, GA, USA

John B. Loomis Colorado State University, Fort Collins, CO, USA

George R. Parsons University of Delaware, Newark, DE, USA

Randall S. Rosenberger Oregon State University, Corvallis, OR, USA

Kathleen Segerson University of Connecticut, Storrs, CT, USA

Laura O. Taylor North Carolina State University, Raleigh, NC, USA

Acronyms

CE/CEs	Choice experiment(s)
EA	Equivalency analysis
GIS	Geographical information system
i.i.d.	Independent and identically distributed
RUM	Random utility model
TCM	Travel-cost model
VSL	Value of statistical life
WTA	Willingness to accept
WTP	Willingness to pay

Use only in Equations

ACS	American community survey
MLS	Multiple listing service
NB	Net benefits
PIN	Parcel identification number
RDC	Research data center
TC	Transaction costs

Chapter 1
Valuing Environmental Goods and Services: An Economic Perspective

Kathleen Segerson

Abstract Nonmarket valuation, i.e., valuing environmental goods and services that are not traded in a market, has been increasingly used in a variety of policy and decision-making contexts. This is one (but not the only) way that researchers and practitioners have sought to define and measure the values that individuals assign to environmental goods and services. The idea of putting a dollar value on protecting the environment has been controversial, but often because the economic approach to valuation has not been well-understood. This chapter provides a nontechnical overview of and rationale for the economic approach to valuation, starting from a broad conceptualization of values versus valuation. It summarizes the economic concept of value and its key features. It then discusses the use of economic valuation in decision making, followed by an overview of the steps involved in the valuation process and important issues that arise in implementing that process. Finally, it identifies and briefly summarizes the principal non-market valuation methods used by economists. In doing so, it sets the stage for the more detailed chapters on theory and methods that follow.

Keywords Preferences · Market failure · Externalities · Ecosystem services · Held versus assigned values · Substitutability · Economic versus commercial values · Economic impacts versus values · Valuation process · Aggregation · Discounting · Uncertainty · Valuation methods

1.1 Making Choices

As Jean-Paul Sartre put it, "we are our choices." Choice is a fundamental part of our lives. We are constantly making choices, often individually or among friends but also collectively. Some individual choices are routine (e.g., about how to spend our income or time on a given day), but others involve major decisions (e.g., about

K. Segerson (✉)
University of Connecticut, Storrs, CT, USA
e-mail: kathleen.segerson@uconn.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_1

houses, families, jobs, or careers). We make collective choices about, for example, the establishment of laws and regulations or the use of publicly owned resources. These collective decisions can be made directly through collective choice mechanisms such as voting or through elected or appointed representatives who make those choices on our behalf.

All choices, whether individual or collective, involve evaluating alternatives so that a choice among those alternatives can be made. Imagine, for example, that you have three hours of uncommitted time and you are trying to decide how to spend it. Suppose you narrow your options down to two: going hiking in a nearby forested area or going to a museum. Assuming that the cost of both alternatives (including travel cost and any entry fee) is the same (say $20), your choice will presumably hinge on which option you prefer. An alternative way to think about this same choice is to ask yourself whether, given $20, would you choose to spend it going hiking, or would you rather keep the $20 and maintain the option of using it for something else, such as going to the museum? Either way of framing the choice—choosing between two activities of equal cost, or choosing between spending the money on hiking or keeping it for some other use—highlights the inherent trade-off involved in nearly all choices, i.e., the fact that choosing one alternative means giving up the other(s).

We can think about collective choice in a similar way, although the choice problem is more complex. Imagine, for example, that $20 million in tax revenue is available for use either to preserve a forested area for hiking or to build a museum. Which option is preferred? Similarly, we could ask whether preserving the forested area is collectively *worth* the $20 million it would cost, or whether instead the money should be used for an alternative, such as the museum. Again, the decision involves a trade-off because using the money for one option means giving up other option(s). Based solely on their own preferences, some individuals might prefer the forested area while others might prefer the museum, so neither is likely to be the preferred choice for all individuals. Thus, collective choice requires not only a comparison and evaluation of the options from the perspective of affected individuals, but also some means of combining disparate individual views into a single collective choice.

Because making choices requires assessing preferences over different options, observing people's choices can reveal information about their preferences. For example, if you face the above choice and choose to go hiking in the forested area, then presumably this implies that you felt that the hiking experience was "worth" the $20 it cost you. Equivalently, through that choice you have revealed that you prefer the hiking to the alternative that the $20 could have bought (a trip to the museum).

The topic of this book, nonmarket valuation, is fundamentally about individual choices and the preferences that underlie those choices. The methods described in the following chapters are all based on the premise that, when faced with a set of options, individuals evaluate those options based on their preferences (and other circumstances) and choose the option that is most preferred, recognizing that choosing one option (e.g., hiking) precludes the other options (e.g., visiting the museum). The information that is obtained about individual preferences can then be used for a variety of purposes, such as informing collective decisions about similar options. For example, information about individuals' preferences regarding use of forested areas can be used by policymakers when making decisions about devoting public funds to preserve those areas or to evaluate the loss that individuals would experience if the forested area were damaged or destroyed.

While nonmarket valuation is fundamentally about individual choices, it is focused on a particular type of choices, namely, those that are not fully captured by purchases or sales in a market. Many choices involving environmental goods and services, including natural amenities such as wilderness and open space, fall into this category. When individuals directly purchase goods and services (such as food or cars), their purchase decisions directly reveal information about their preferences for these items. However, individuals do not typically make a direct purchase of environmental goods and services such as clean air or clean water (although they might purchase a trip to a wilderness area). For goods that are not directly for sale in the market, individuals cannot express their preferences through their purchases. The nonmarket valuation methods described in this book are designed to elicit information about those preferences through other means.

1.2 Choices and Market Failure

When individuals make choices based on their preferences and self-interest, the outcomes that result can be good for society as a whole as well. This is the essence of Adam Smith's observation in "The Wealth of Nations," published in 1776, that individuals are led by an "invisible hand" to unintentionally promote broader social goals. However, in many cases, the invisible hand does not work, i.e., individual decisions based solely on self-interest do not lead to outcomes that are best for society. The invisible-hand argument rests on the assumption that markets exist for all the goods and services that individuals care about, thereby creating a means for buyers to express their preferences in the marketplace. However, as noted above, for many environmental goods or services, markets do not exist.

The lack of markets for many environmental (and other nonmarket) goods has important implications for resource allocation.[1] In particular, a purely market-based economy will tend to underprovide nonmarket goods relative to what would be socially optimal.

We can think about the problem in two alternative (but equivalent) ways. The first views environmental improvements as goods or services that would be supplied by individuals or firms, if only a market existed. For marketed goods and

[1]The reasons that markets do not exist can vary. In many cases, the market failure arises because the environmental good or service is a public good. For example, air quality in a city is a pure public good because all individuals living in the city will benefit from an improvement in the city's air quality and no one in the city can be excluded from enjoying the benefits of the improvement. Public goods suffer from the "free-rider" problem, which can impede the development of a market for the good. Markets can fail to exist for other reasons as well, including ill-defined property rights, information asymmetries, and difficulty in defining and monitoring tradable units.

services (e.g., food and housing), suppliers provide these goods in exchange for payments that cover their costs of providing them. In contrast, when there is no market for an environmental good, individuals or firms who could supply that good will not have an incentive to do so because they will not receive a payment to cover the associated costs. For example, private landowners will not have an incentive to keep some of their land as protected habitat if they have no way of recouping the foregone profits by selling the environmental services that would result. So, unlike with marketed goods, even if the benefits (to society) from providing those services exceed the corresponding cost (i.e., the foregone profit), the landowner is not likely to supply the protected habitat.

The second way to think about the undersupply of environmental goods is based on the concept of externalities, i.e., unintended (and uncompensated) positive or negative impacts that one individual's or firm's decisions have on others. Environmental degradation is a classic example of a negative externality. By engaging in activities that, as a byproduct, degrade the environment, individuals or firms impose environmental damages on others. When no market exists in which those individuals must purchase the right to impose those damages, the individuals will face only their own private costs of engaging in the activity rather than the full social cost (which includes the environmental cost). As a result, they will tend to overengage in the environmentally degrading activity. For example, an electric utility that generates carbon dioxide emissions as a byproduct of electricity production will pay for the labor, capital, and fuel it uses in that production, but it will not typically pay for the cost of the pollution it generates. Because emissions are not costly to the utility, it has no incentive to try to reduce its emissions and, therefore, will typically pollute too much (i.e., undersupply environmental protection).

The undersupply that results from missing markets creates the opportunity to improve outcomes from society's perspective by (1) facilitating the creation of those markets, if possible, or (2) seeking to provide the good through means other than markets, such as through laws/regulations requiring or restricting certain activities or through direct provision by the government. In either case, information about the value of the environmental goods and services that would be supplied can help in addressing and overcoming the market failure, and nonmarket valuation can play a key role in providing that information. Consequently, the need for nonmarket valuation often arises in the context of missing markets or market failure.

1.3 Development of Nonmarket Valuation

Most nonmarket valuation techniques first appeared in the U.S. in the 1950s, primarily for use by federal agencies in benefit-cost analyses of proposed water resource projects such as dam construction. In the years that followed, environmental and natural resource economists refined and improved these techniques and applied them in a wide variety of contexts. Progress was spurred on in the early 1980s with two federal actions. One was Executive Order 12291 (issued in 1981), requiring benefit-cost analyses of all proposed major regulations (see Smith 1984). The other was passage of the Comprehensive Environmental Response, Compensation and Liability Act (passed in 1980), requiring an assessment of damages to natural resources from releases and spills (Kopp and Smith 1993; Portney 1994). These and subsequent actions in the U.S. and elsewhere focusing on public land management and environmental protection led to many applications of nonmarket valuation methods, primarily to assess the environmental and health benefits of environmental regulation, to estimate compensation for damages suffered as a result of spills or other types of contamination, and to inform land and water management decisions (see, for example, Smith 1993; Adamowicz 2004; Carson 2012).

Interest in the use of nonmarket valuation techniques among non-economists is more recent and stems to a large extent from the growing understanding that the natural environment generates "ecosystem services" that sustain and enhance human well-being and the recognition that those services are being significantly degraded or threatened by a wide variety of activities across the globe (Daily 1997; Millennium Ecosystem Assessment 2005). In addition, ecologists realized that these critical services were being given little, if any, weight in policy decisions because their contributions to individual and collective well-being were not being estimated and included along with other considerations in evaluating choices. This led to increased interest in valuing ecosystem services and including those values in decision-making (National Research Council 2005; Carpenter et al. 2006; Brown et al. 2007).

An early attempt to place a monetary value on the contributions of the world's ecosystems estimated the mean annual value to be $33 trillion (Costanza et al. 1997), suggesting that global ecosystem services were "worth" more than the annual global production of marketed goods and services at that time. While the methods and results used in this analysis were heavily criticized by economists (see, e.g., Toman 1998; Bockstael et al. 2000), this early work and the discussion it spurred highlighted the importance of considering ecosystem services in individual and collective decisions and the role that nonmarket valuation techniques could play in ensuring that consideration. It also highlighted the need to understand and apply those methods appropriately. This book is designed to meet the growing demand for the use of nonmarket valuation techniques and to provide the necessary foundation for understanding and appropriately applying those techniques.

1.4 Values Versus Valuation

Nonmarket valuation is often described as a means of "valuing" the environment (or environmental goods and services). However, the concept of "value" or a person's "values" encompasses a wide range of ideas, and there is often confusion over what exactly is meant when we refer to valuing something (see Brown 1984; Dietz et al. 2005). For example, individuals can value certain types of behavior (such as loyalty), certain end states (such as freedom), and certain qualities (such as beauty). Brown (1984) refers to these end states and other ideas of what is good or preferable as *held* values. In contrast, he refers to the values that individuals place on an object as *assigned* values, which, importantly, are "not a characteristic of the object itself but rather the standing of the object relative to other objects" (Brown 1984, p. 233). The value that an individual assigns to an object (relative to other objects) will depend on a number of factors, including "(1) the person's perception of the object and all other relevant objects, (2) the person's held values and associated preferences, and (3) the context of the valuation" (Brown 1984, p. 235), where context is broadly defined to include the external and internal circumstances of the valuator, the way in which values are expressed, and whose interests the valuator is representing (e.g., pure self-interest or a broader constituency).

The distinction between held values and assigned values is critical in understanding nonmarket valuation as a means of "valuing" the environment (or environmental goods and services). Individuals may have held values related to environmental protection, i.e., they may feel that environmental protection is an important and desirable type of behavior or end state. These values can be based on a number of possible grounds, such as spirituality, bioethics, or aesthetics (e.g., beauty). However, they are not by themselves directly measurable in economic terms, so they are not the focus of nonmarket valuation.

Nonmarket valuation seeks to measure assigned values, which are influenced by held values but distinct from them. Rather than seeking to value environmental protection as a general principle, it seeks to measure the value that individuals assign to particular environmental quality (or natural resource) outcomes *relative to* some alternative. For example, it is focused on the value an individual would assign to having air quality at level A instead of having air quality at some alternative level, say B. In this case, the *object* to which the value is assigned is the *change* in air quality (from A to B). Thus, the values measured by nonmarket valuation are always relative in the sense of being assigned to *changes* from one outcome or scenario to another.

It is important to note that nonmarket valuation does not seek to identify (let alone measure) the underlying held values that are manifested in the assigned value for a given change. For example, it does not seek to identify whether an individual values an improvement in air quality or preservation of a wilderness area based on spiritual, bioethical, aesthetic, or some other grounds. In other words, it does not seek to identify, understand, judge, or explain the *reason* that an individual assigns a particular value to the change and does not involve a process designed to influence the underlying held values (Polasky and Segerson 2009). Regardless of the underlying philosophical basis or reason, nonmarket valuation simply seeks to measure the values that individuals assign to a given change based on their preferences over alternative outcomes and the trade-offs they are willing to make. In brief, it seeks to measure what changes people care about and how much they care, independent of *why* they care.

Although held values can be stated in terms of general principles (e.g., "I value my health"), assigned values must be stated in terms of some scale that allows a direct comparison to determine whether one object or change is valued more, less, or the same as another. The term "valuation" refers to the process of measuring individuals' assigned values using a given scale. Fundamentally, this process involves two primary components: determination of the relevant change(s) to be valued and estimation of the value of the change(s) based on a given scale.[2] Different scales or means of expressing assigned values exist (Brown 1984), and there are different views on the importance of the factors that influence those values. These views tend to vary across scholarly disciplines. For example, economists generally emphasize the importance of (fixed) preferences and income, while psychologists and sociologists focus on other internal and external factors, such as perceptions, social/cultural influences, and framing effects. As a result, different disciplines tend to view valuation somewhat differently, employ different methods for eliciting assigned values, and express those values using different measures or scales (Dietz et al. 2005; U.S. Environmental Protection Agency 2009).

This book focuses on an economic approach to valuation. As mentioned, it is one (but not the only) way that researchers and practitioners have sought to define and measure the values that individuals assign to environmental goods and services.[3] Although economic valuation does not necessarily capture all relevant dimensions of assigned value in a given context and may not be appropriate in all circumstances, it is based on a well-developed theoretical foundation and has proven to be very useful in practice as a means of ensuring that the environmental or health impacts (either positive or negative) of individual or collective choices are considered when those choices are made.

1.5 The Economic Concept of Value

Standard economic theory defines value in terms of the trade-offs that individuals are willing to make. The value of something, such as an improvement in environmental quality (call this change X), is the maximum amount of something else (call this good Z) that an individual would be willing to give up in exchange for the change that is being valued.[4] This presumes that, for any reduction in the quantity of some good or service, there is an increase in the quantity of some other good or service that would leave the individual at the same level of well-being ("utility") as before.

Two fundamental implications of this definition are: (1) the more of good Z that an individual is willing to give up to get X, the more the individual values X; and

[2]Most theoretical discussions of economic values and nonmarket valuation methods focus on valuing a single change (e.g., a change in ambient air quality). However, most real-world valuation contexts involve changes in multiple environmental goods or services and multiple impacts on human well-being. The need to consider multiple changes or impacts raises questions about interconnectedness and aggregation and clearly complicates the valuation process. See National Research Council (2005, Chapter 5) for a useful discussion of valuing changes in multiple ecosystem services.

[3]An important research question is whether alternative ways to define and measure assigned values yield consistent information about underlying preferences. The limited evidence that exists on this is mixed. See, for example, Cooper et al. (2004) and Spash (2006).

[4]See Chap. 2 for a more formal definition.

(2) if the maximum amount of Z the individual is willing to give up to get X is greater than the maximum amount he or she is willing to give up to get Y, then the individual values X more than Y. The scale used to measure (and compare) values is therefore the maximum amount of Z the individual would be willing to give up. Note that nothing in this definition of value precludes X from being something to which the individual assigns a negative value. For example, if X represents a given amount of environmental degradation, then the amount of some beneficial good Z that the individual would we willing to give up to get X would be a negative number, meaning that the individual would actually require compensation (i.e., more Z) to accept X.

This concept of value does not require that values be expressed in monetary terms, i.e., that Z be money. Any other good that individuals care about could be the basis for the expression of economic values. For example, the economic value of reducing one type of risk (such as fatality risk from natural disasters) can be expressed in terms of the increase in another type of risk (such as fatality risk from traffic accidents) that the individual would be willing to accept (Viscusi 2009). Values expressed in these terms are generally called "risk-risk trade-offs."

In principle, Z can be anything individuals care about, but in practice, economists typically seek to measure values in monetary terms, i.e., Z is taken to be an amount of money an individual would be willing to give up to get X (i.e., the individual's "willingness to pay" [WTP] for X) or the amount of monetary compensation he would require to give up X (i.e., the individual's "willingness to accept" [WTA] compensation for not getting X).[5] When X represents something the individual views as beneficial (i.e., he or she would prefer having it to not having it and so would prefer not to give it up) and Z is money, economists refer to the assigned monetary value as the *benefit* of X, representing what X is *worth* to the individual. Thus, while in everyday language the word "benefit" is broadly used to refer to something beneficial (e.g., a benefit of improved air quality is reduced infant mortality), in the context of economic valuation, the term "benefit" has a much more specific meaning based on the economic concept of assigned value, a meaning that is not only quantitative but monetary as well. Having values (benefits) expressed in monetary terms allows for a simple means of aggregating values across individuals and comparing them to costs.

The economic concept of value reflects four key features:

---

[5]Because people regularly use money to buy things and accept money when they sell things, the idea of trading money for goods and services is familiar to them (although they may never have traded for the specific good or service being valued). Nonetheless, some individuals may feel that certain things, such as changes in health or environmental quality, should not be "commodified" and sold in markets. This does not, however, mean that providing those things does not involve trade-offs. As emphasized, economic values are fundamentally about the trade-offs individuals would be *willing* to make. Even individuals who object to the idea of buying or selling nature exhibit a willingness to make trade-offs related to health and the environment in their everyday lives (for example, every time they travel by car).

1. The values that individuals assign depend on their preferences over different outcomes, which are assumed to be stable and consistent (in the sense of not being unduly influenced by issues such as framing, presentation, or elicitation method).[6] Because individuals assign the values, they are anthropogenic, i.e., they are derived from humans and are not defined independently of the individuals who assign them.
2. Although economic values are agnostic on the reason(s) individuals care about and hence value something, they do assume there is some (finite) substitutability between what is being valued and other things the individual cares about. In other words, they assume that individuals care about multiple things (such as environmental quality, health, food, and leisure time) and are willing to make trade-offs among these, at least over certain ranges. Held values that are absolute and do not allow for *any* substitutability (e.g., "freedom at all costs") preclude measurement of assigned values in economic terms.
3. As noted, values are assigned to *changes*. These changes can be purely hypothetical or actual realized or predicted changes. They can be expressed in absolute terms (e.g., 100 additional acres of wetlands), percentage changes (e.g., a 10% reduction in ambient concentration of particulates), or as a with-or-without scenario (e.g., with or without an old-growth forest area). However, the changes must be feasible. This implies that when using a valuation technique that involves asking people about hypothetical changes, the changes must be meaningful to the individuals asked to value them. For example, asking individuals to assign values to the entire global ecosystem is not meaningful because it requires that individuals envision the world without that ecosystem, which is probably an impossible task, in part because it is an impossible change.
4. In general, economic values will depend not only on preferences but also on how much Z an individual has available to trade. When measured in monetary terms, this means that values (benefits) depend on an individual's income. This feature is not problematic when comparing values for a given individual. For example, if an individual is willing to pay more of his income to get X than to get Y, then presumably he values X more than Y, regardless of the amount of income he has. However, comparisons are less clear when made across individuals because two individuals with identical preferences but different incomes could express different values for the same X. As a result, if a wealthy person expresses a higher willingness to pay for X than a poor person, this would imply that the benefit of X as defined by economic value is greater for the wealthy person than the poor person, even though it does not in a broader sense imply that X is more important to the wealthy person or that the wealthy person cares more about X than does the poor person (see Sect. 1.8.2 for further discussion).

Although economic values are typically expressed in monetary terms, as already mentioned, economic valuation is *not* limited to goods and services that are bought

[6]In contrast, some psychologists believe that preferences are *constructed* through the elicitation process. See, for example, Lichtenstein and Slovic (2006).

and sold in markets. In fact, the purpose of nonmarket valuation is to elicit information about the values individuals would assign to things that are *not* bought and sold in markets. As long as an individual cares about something, regardless of whether it can be bought and sold and regardless of the reason the individual cares, he or she will presumably assign a nonzero value to it. Consequently, the economic concept of value is fundamentally different from and should not be confused with the concept of commercial value. While commercial values often reflect economic values, very frequently they do not. In particular, goods that are not sold in markets typically have no commercial value even though the economic value individuals assign to them can be very large. For example, people might assign very high values to improvements in their health or reductions in their risk of getting cancer even though these cannot be directly purchased in a market. Similarly, they might assign very high values to an increase in biodiversity even though biodiversity cannot generally be directly bought or sold for money. Therefore, economic values reflect a much broader notion of value than commercial values.

Figure 1.1 illustrates a standard classification of economic values that highlights the breadth of what the concept covers (National Research Council 2005). In particular, it shows that the total economic value of a natural resource or environmental good includes not only the benefits individuals get through use of the good (use values) but also the value they place on the good even if they do not actually use or come in contact with it (passive-use or nonuse values). The latter values arise when an individual values the *existence* of a species or preservation of a natural environment for reasons such as bioethics, cultural heritage, or altruism toward others (including future generations). For example, empirical evidence suggests that passive-use values exist for the protection of not only charismatic species (such as grizzly bears and bighorn sheep; see Brookshire et al. 1983) but also marine habitats (McVittie and Moran 2010) and even moss (Cerda et al. 2013).[7]

Use values arise as a result of use of or physical contact with the environmental good. Use can be either consumptive or nonconsumptive. Consumptive use implies that use by one person precludes use by another. Examples include the harvesting (and use) of timber or the taking of a pheasant while hunting. With nonconsumptive use, one individual's use does not diminish the amount available for others. For example, bird watching by one individual does not diminish the potential for others to enjoy the same viewing, and swimming or boating in a lake by one individual does not preclude others from enjoying the same recreational experience. These are examples of direct nonconsumptive use, where the individual directly benefits from the environmental good or service. Nonconsumptive use can also be indirect, such as when we benefit from climate regulation provided by forests or storm protection provided by wetlands.

[7]However, see Common et al. (1997) for evidence suggesting that the assumption of substitutability that underlies the economic notion of existence value may not hold in all contexts.

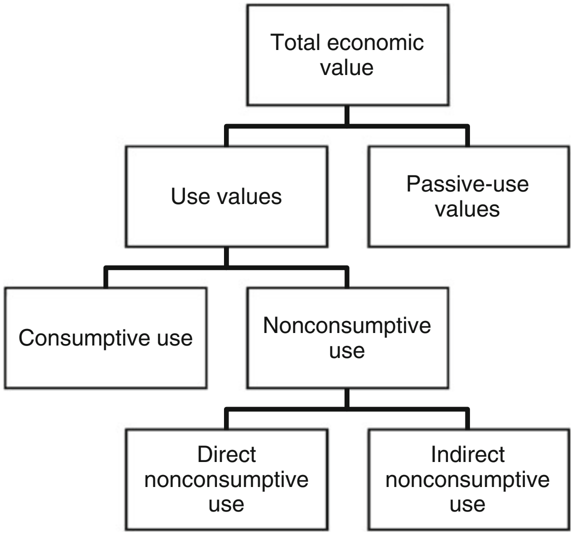


Fig. 1.1 Classification of economic values

Thinking about economic values using a categorization like the one illustrated in Fig. 1.1 helps in recognizing the types of values that are included and in ensuring that some types of values are not overlooked. In addition, it helps to avoid double counting values, which would lead to an overestimation of total economic value. Suppose a resource manager seeks to estimate the benefits of a program designed to improve the marine habitat that supports a commercial fishery. It would be double counting to include in a measure of total economic value both the consumptive-use value of the increased fish catch and the indirect-use value from the habitat improvement that in turn allows the increased catch because the two are just different manifestations of the same benefit. A careful delineation of the types of benefits using a categorization such as the one in Fig. 1.1 helps to ensure that all components of total economic value are included, but each is included only once.

The above discussion focuses on what the concept of economic value includes. It is equally important to recognize the factors that are *not* valid components of economic value. In particular, it is important not to confuse the concept of economic value (benefits) with the concept of economic impacts. Consider again a program designed to improve marine habitat that supports a commercial fishery. If the program is successful in improving fish catch, it could lead to an increase in economic activity in the region, which could in turn generate additional jobs and income. While the impact on regional employment and income could be an important consideration when evaluating the program, these impacts are *not* measures of the economic benefit of the program. More specifically, they are not a measure of the amount individuals would be willing to pay for the program or the amount they would require in compensation to give it up.

As an illustration of why impacts do not measure benefits, consider the impact of an oil spill that threatens to destroy marine habitat. Cleanup of the spill can generate jobs and income, but clearly this does not mean the spill was beneficial. In fact, the labor needed to clean up the spill is part of the cost of the spill, not a benefit of the spill. As a second example, consider a regulation that restricts the type of gear that can be used in a fishery in order to reduce bycatch of a protected species. If, as a result of the regulation, it now takes more labor to catch a given amount of fish, employment in the industry may increase. However, this increased labor requirement is a measure of the cost of the regulation, not the benefit of the regulation. The regulation is costly because, among other things, it now takes more effort to produce the same amount of harvest. The benefit of the regulation would come from the value of the species protection, not from the employment impacts of the regulation.[8] These examples illustrate why, although economic impacts might be an important consideration in evaluating alternatives, they should not be confused with economic benefits.

1.6 Use of Economic Values

Information about the economic values that individuals assign to environmental changes can help improve decisions in a wide variety of contexts (e.g., Adamowicz 2004). These include decisions about (1) public policies (at the local, regional, or national level), (2) resource allocation and priorities, (3) compensation for losses, and (4) design of environmental markets.

When making public policy decisions regarding, for example, laws and regulations, policymakers can use a number of different criteria or decision rules. One possibility is to base decisions on an explicit comparison of benefits and costs.[9] Of course, the use of a benefit-cost criterion requires a monetary measure of benefits. However, even if a strict benefit-cost criterion is not used, information about benefits (and costs) can be very helpful in evaluating alternatives (Arrow et al. 1996). Even if decisions are based on sustainability or precautionary principles, information about benefits can help in identifying the trade-offs implied by those decisions, especially when they involve conflicts regarding different environmental goals or ecosystem services (e.g., timber production versus carbon sequestration).

Similarly, resource managers who need to allocate fixed budgets across different projects, programs, or initiatives can use information about economic values to ensure that resources are targeted in a way that maximizes benefits. For example, in making

[8]Under some conditions, it is possible to value a change in output by aggregating the payments to the owners of the inputs used to produce the additional output. For example, with perfectly competitive markets and constant returns to scale, the value of additional agricultural output can be measured as the sum of the payments to the laborers, landowner, fertilizer producers, etc., who supply the inputs used to increase production. In this case, the income paid to farm workers can be used as part of the measure of benefits. However, including both the payments to the inputs and the revenue from the sale of the additional output as measures of benefits would be double counting. Benefits can be measured as the value of the inputs or the value of the output but not both.

[9]The theoretical foundations of benefit-cost analysis as a basis for public policy are well established and described in detail in Just et al. (2004).

land purchases to protect open space or habitat, the available funds can be allocated to the purchase of parcels that give the greatest benefit per dollar spent. Identifying those parcels requires information about the benefits that preservation of specific parcels would generate based on their environmental and other characteristics.

Measures of the economic value of environmental losses can be used in damage assessments and in the determination of the amount of monetary compensation that would be necessary to restore the well-being of affected individuals to their pre-loss levels. Information about benefits can be useful even if compensation is in-kind rather than monetary. For example, if wetland losses in one location are offset by wetland restoration elsewhere, measures of the benefits generated by wetlands in the two locations can be used to determine the amount of restored area needed to compensate for the loss in well-being resulting from the original reduction in wetland area.

In addition to policymakers, government officials and resource managers, private parties and nongovernmental organizations (NGOs) might also want to use information about economic values in setting priorities and designing initiatives. For example, there is increasing interest in the development of markets for environmental goods and services such as ecotourism, carbon sequestration, and habitat preservation for biodiversity (see Daily and Ellison 2002; Heal 2000; Brown et al. 2007). Information about the benefits associated with these services can be used to determine payments or contributions that individuals would be willing to make for the purchase of these services. In addition, it can be used to target specific groups for contributions or to design conservation and other programs to align with preferences of specific constituencies, such as donors.

1.7 The Valuation Process

Although economic values may be used in various decision contexts, in all cases the basic process for environmental valuation is essentially the same. The key steps are listed in Table 1.1 (adapted from U.S. Environmental Protection Agency 2009).

As an example, consider the process of estimating the benefits from a restoration project that would increase or restore stream flow and/or fish passage in a river with a dam.[10] Step 1 would identify the possible alternatives under consideration (such as full removal of the dam, partial removal of the dam, installation of fish ladders, etc.). Each alternative could generate a number of biophysical changes (identified in Step 2) that might be important to individuals either because of use value (for example, increased fish abundance increases catch rates in recreational or commercial fishing) or because of passive-use values (for example, individuals may value the existence of a more viable population of native fish in the river or the wildlife that depend on those fish for food). Once these potential sources of value have been identified (Step 3), the relevant impacts can be quantified (Step 4) and valued (Step 5).

[10]For an example of nonmarket valuation in this context, see Johnston et al. (2011).

Table 1.1 Steps in the valuation process

Step 1	Identify the decisions that need to be made and the options to be considered. This step is often referred to as "problem formulation"
Step 2	Identify the significant environmental or biophysical changes that could result from the different options
Step 3	Identify the types of impacts these biophysical changes might have on human well-being and so could be important to individuals
Step 4	Predict or hypothesize the quantitative magnitude of environmental changes in biophysical terms that are relevant to human well-being and hence can be valued
Step 5	Estimate the economic values that individuals would assign to these changes using appropriate valuation methods
Step 6	Communicate the results to the relevant decision-makers

Source U.S. Environmental Protection Agency (2009)

Although Table 1.1 depicts the valuation process as linear, in fact it is an iterative process because information generated by later steps might imply a need to revisit some previous steps. For example, in the process of estimating the values individuals assign to the various impacts, some unanticipated sources of value might be discovered. In the above example, it might be discovered in Step 5 (perhaps through survey responses or focus groups) that in addition to the relevant impacts already identified, individuals value the overall ecological condition of the river, which would be affected by the restoration project. This would necessitate going back to Step 4 to quantify the resulting change in ecological condition in units or measures that relate directly to what individuals value and then revisiting Step 5 to estimate this additional value.

This book focuses, of course, on methods that can be used in Step 5. However, it should be clear from the discussion above that Step 5 is part of a broader valuation process that requires a close collaboration between natural scientists and social scientists. For example, an interdisciplinary team should be involved in Steps 1-3 to provide the different perspectives and expertise needed to ensure that the remainder of the valuation process (Steps 4-6) is focused on the impacts that are most important in terms of their contribution to human well-being. Otherwise, natural scientists can end up predicting biophysical changes in terms that cannot be readily valued by individuals (such as impacts on phytoplankton), or social scientists can end up valuing changes that are not closely related to the biophysical impacts of the alternatives that are being considered.

1.8 Some Additional Issues

In moving from conceptualizing the valuation process described above to actually estimating values (benefits) in Step 5 using nonmarket valuation, a number of issues can arise. These include: (1) whose values to include, (2) how to aggregate across individuals, (3) how to aggregate across time, and (4) how to treat uncertainty.[11]

[11]For more detailed discussions of these and related issues, see, for example, Freeman et al. (2014).

1.8.1 Whose Values to Include

Critical in determining how much individuals value a particular change in environmental goods and services is defining the relevant population of individuals. While the answer to this might seem to be that anyone who values the change should be included, the relevant population actually depends on the valuation context. For example, in a context where valuation is designed to estimate the compensation to pay to individuals who are harmed by environmental degradation from an oil spill, the relevant population is typically the set of individuals who are impacted *and* legally entitled to compensation. If the damage payment is designed to compensate the public for kills of wildlife species that have existence value (such as birds or seals), the question is how the "public" is defined, i.e., whether it includes the existence values across *all* individuals (i.e., the global population) or some more locally defined public. The amount of compensation would clearly be greater in the former case than in the latter.

Similarly, in evaluating the costs and benefits of a national policy to reduce greenhouse gas emissions and slow global climate change, should the benefits that are included be just the benefits realized by individuals within that country, or should benefits be measured at a global scale? Again, the measure of benefits will be much higher if it includes benefits to people everywhere and not just those within the country considering the policy. Whether a global or more local measure of benefits is appropriate depends on how the policy decision will be made. For example, if policymakers are willing to adopt the policy as long as global benefits exceed the costs that the country would incur (even if local benefits do not), then the benefit measure should be at the global scale. However, if policymakers will base their decision on whether their country will realize a net benefit, i.e., the benefits within the country exceed the costs, a more localized measure of benefits is needed.

1.8.2 Aggregating Values Across Individuals

When the relevant population has been identified, estimating total value across that population requires a means of aggregating values across individuals. Individual values measured in monetary terms are typically aggregated by simply adding these values over the relevant individuals.[12] This sum is typically unweighted, implying that the values of all individuals receive the same weight in the calculation of the

[12]As an alternative, preferences could be aggregated across individuals simply by counting the number of individuals who prefer one option to another. This approach underlies decision rules based on standard voting procedures. A key drawback to using votes as a means of aggregating preferences to determine outcomes is that the resulting decisions do not reflect the intensity of individual preferences for one option over another. For example, in a three-person vote, an option that is only slightly preferred by two individuals would win over an option that is strongly preferred by the third individual.

total, regardless of the characteristics or circumstances of those individuals.[13] The aggregate benefit measures used in benefit-cost analyses use this approach,[14] where the objective is to identify options that generate the greatest good for the greatest number.[15]

While equally weighting benefits across all individuals may seem fair in the sense of implying equal treatment for all, it is important to understand its implications. Recall that in general for any good or service, including environmental goods, the value an individual assigns to an increase in that good (for example, the amount the individual is willing to pay for that increase) will depend on his or her income. This implies that, for two individuals with identical preferences, in general the one with the higher income will place a higher value on (have a greater willingness to pay for) an increase in the good than the person with the lower income. In other words, when all else is equal, benefits will typically be higher for wealthier people. This implies that equally weighting benefits does not actually give equal weight to the preferences of all individuals.

To see this, consider a change (call it X) in an exclusive good that is valued by two individuals with identical preferences but different incomes. Assume Person 1 has the higher income and the income difference causes Person 1 to place a higher economic value on X than Person 2 does. Assume Person 1 is willing to pay $100 for X, while, because of lower income, Person 2 is only willing to pay $50. As a result, the good would be viewed as generating greater benefits if it were consumed by Person 1 than if it were consumed by Person 2. More generally, with equal weighting and when all else is equal, options that generate value for the wealthy

[13]Occasionally, a weighted sum will be used in an effort to incorporate distributional concerns, i.e., to give more weight to the benefits that accrue to one group of individuals than to another (see, for example, Johannsen-Stenman 2005). However, most economists do not advocate the use of a weighted average as a means to incorporate distributional concerns; rather, they advocate providing decision-makers with an unweighted measure of total benefits along with information about the distribution of benefits across relevant subpopulations. See, for example, Arrow et al. (1996).

[14]This is based on the compensation principle that underlies benefit-cost analysis. For a detailed discussion, see Just et al. (2004).

[15]Note, however, that maximizing aggregate net benefits or aggregate income is not generally equivalent to maximizing the sum of utility across all individuals. The two will be the same *if* coupled with a redistribution of income that equates the marginal utility of income across all individuals. Absent that redistribution, a change for which benefits exceed costs could actually reduce aggregate utility. To see this, consider a choice x, where an increase in x generates an increase in income (i.e., a benefit) for Person 1 and a decrease in income (i.e., a cost) for Person 2. Let $Y_i(x)$ be income for Person i (where $i = 1$ or 2) and let $u_i(Y_i(x))$ be i's utility. Then, if $\frac{\partial u_1}{\partial Y_1} < \frac{\partial u_2}{\partial Y_2}$ and there is no actual compensation, it is possible to have $\frac{\partial Y_1}{\partial x} + \frac{\partial Y_2}{\partial x} > 0$ (i.e., the gain to person 1 exceeds the loss to person 2) even though $\frac{\partial u_1}{\partial Y_1}\frac{\partial Y_1}{\partial x} + \frac{\partial u_2}{\partial Y_2}\frac{\partial Y_2}{\partial x} < 0$ (i.e., aggregate utility decreases). While it might be reasonable to assume that direct mechanisms (such as taxes) are available to address income distribution within a given generation, this assumption is much more questionable for income redistribution across generations. For this reason, differences in the marginal utility of income across individuals are often not explicitly considered when aggregating benefits and costs at a given point in time, while differences in the marginal utility of consumption across generations play an important role in aggregation across time (see discussion in Sect. 1.8.3).

will yield higher economic benefits than those that generate value for low-income groups. As such, although in nonmarket valuation it is standard practice to use an unweighted sum of benefits across individuals to measure the aggregate benefit across the relevant population, the implications of this approach need to be borne in mind when interpreting these measures.

1.8.3 Aggregating Across Time

In many contexts, such as climate change, the benefits of a particular policy or action extend across many years. A measure of the total benefit of such an action should include not only benefits in the current period but also future benefits that would result from the current action. In principle, it is possible to think about the total benefit to an individual simply as the total amount the individual would be willing to pay today for the policy change, recognizing the stream of impacts it would have over time.[16] In practice, however, benefits are typically measured separately for each time period and then aggregated over time. Accordingly, when benefits extend across time, estimating total benefits typically requires some means of aggregating over time. The standard approach used in economics is to weight benefits that occur at different points in time using a discount factor and then add the weighted measures of benefits across all time periods to get a measure of the total (discounted) benefit. Usually this is done using a constant discount rate, defined to be the rate at which the weights change over time. However, uncertainty about the appropriate discount rate to use can provide a rationale for use of a discount rate that declines over time.[17]

Discounting future benefits can be controversial and has important implications for measures of total benefits, making it important to understand the economic rationale for discounting. One rationale stems simply from the ability to earn a return on investments. As an example, if you can earn interest at a rate of 5% on your investments, then you should be indifferent between receiving a payment of $100 today (and investing it, so that you have $105 in the next period) and receiving a payment of $105 in the future period. Similarly, if you instead receive $100 in the next period, it is worth less to you than if you had received $100 today because of the lost investment opportunity. This means that payments received in the future should be discounted when compared to payments received today to

---

[16]However, to the extent that current decisions affect future generations, the individuals in those generations are not around today to express the values they would assign to the relevant changes. Thus, in practice it is the current generation that must express values on behalf of future generations.

[17]A simple example illustrating this result is provided in Cropper (2012). A declining discount rate is also consistent with some forms of nonstandard preferences, such as those that exhibit hyperbolic or quasi-hyperbolic discounting. See, for example, the discussion and references in Benhabib et al. (2010).

account for this lost opportunity. Of course, this simple example ignores real-world complications associated with investments (such as uncertainty regarding the return on most investments and taxes on investment income), but it illustrates an investment-based rationale for discounting.

There is also a possible consumption-based rationale for discounting, based on either of two considerations about consumption over time. First, individuals may simply value current utility more than future utility and as such put more weight on increases in current consumption than increases in future consumption. Second, if an individual expects his or her income (and, therefore, consumption) to rise over time and the marginal utility of an additional dollar's worth of consumption decreases as income increases, then additional consumption dollars in the future will give less utility than those in the current period simply because of the difference in income. Assuming growth in the individual's income over time, both considerations imply a positive discount rate.

The consumption-based rationale for discounting becomes more complicated, however, when the benefits accrue well into the future, as will occur for decisions with long-term impacts that last for generations (such as those related to climate change). In this case, the issue of aggregating over time is confounded with the issue of aggregating across groups of individuals (generations). Now, the values placed on current versus future utility are no longer simply a reflection of an individual's preferences. Rather, if used in policy decisions, these values reflect society's judgment about whether the well-being of one group (e.g., the current generation) should receive more weight than the well-being of another group (e.g., a future generation). If, as many argue, there is no reason to believe that the well-being of one generation is more important than the well-being of another, then the well-being of all generations should be weighted equally in aggregating well-being across generations (see, for example, Heal 2005, 2009).

However, this does not necessarily imply that benefits to future generations should not be discounted. For example, as with the case where benefits accrue to the same individual over time, if, other things being equal, the marginal utility of consumption diminishes as consumption increases, then benefits should be weighted differently for the different generations, based solely on their consumption levels (not on when they live). If future generations are expected to have a higher consumption level (due to economic growth), then even if the utility of all generations are weighted equally, the discount rate should be positive, implying that benefits to future generations receive less weight than those that accrue to the current generation.[18] This implies that choice of the appropriate discount rate to use in aggregating benefits across generations is not simply a question of *picking* a discount rate; rather, the rate used should reflect a number of considerations, including how society chooses to weight the utilities of different generations (based

[18]A negative discount rate might arise, for example, in the context of ecosystem services where those services are becoming scarcer (rather than more abundant) over time. See, for example, Heal (2005, 2009) and National Research Council (2005).

on intergenerational equity) and information about the rate at which consumption is expected to change over time. Because of the difficulty of determining a single *correct* discount rate, aggregate measures of benefits are often computed for a range of discount rates.

1.8.4 Uncertainty

In many (if not most) cases, valuing environmental changes will involve uncertainty. Uncertainty can arise either in predicting the magnitude of the environmental changes to be valued or in assigning values to those changes.

Consider, for example, a policy to reduce greenhouse gas emissions in an effort to mitigate climate change. Valuing the benefits of such a policy first requires an estimate of the changes that will result. Predicting those changes will involve many sources of uncertainty, including uncertainty about how the policy will affect emissions, how the change in emissions will affect the climate, and how the resulting change in climate (e.g., the distributions of temperature and precipitation at various locations) will affect environmental, health, and other outcomes.

Models are often used to predict these impacts, and there is uncertainty about both the appropriate model structure and the model parameters (National Research Council 2005). In addition, we may be uncertain about our own future circumstances (such as future preferences and income) and/or the preferences of future generations. The values assigned to the predicted changes will reflect these uncertainties as well.

Methods exist to address uncertainty in valuation—both theoretically and in practice. The theory that underlies the economic concept of value described above can be extended to define values in terms of the trade-offs individuals are willing to make *given the uncertainty* associated with the impacts and the factors that affect values, as long as it is possible to identify all possible outcomes and the probability of each occurring. This generalization can incorporate not only uncertainty but also the possibility of learning (and hence reducing uncertainty) over time (Zhao and Kling 2009). Of course, because economic values are defined in terms of trade-offs that would hold an individual's well-being constant, in the presence of uncertainty the values are typically defined in terms of trade-offs that would hold the *expected value* of an individual's well-being ("expected utility") constant. Such measures of economic values reflect not only the individual's preferences over alternative outcomes but also his or her preferences over different amounts of risk, i.e., the extent to which the individual is averse to risk, enjoys risk, or is indifferent to risk. Therefore, while the basic concept of economic value remains the same with or without uncertainty, the values that individuals assign to a given change will reflect additional considerations when uncertainty exists.

In practice, uncertainty about the changes to be valued can be addressed through use of techniques such as sensitivity analysis or Monte Carlo simulation, which can provide information about the distribution of possible outcomes (see National

Research Council 2005). Survey methods can also be used to elicit information about values that explicitly reflects uncertainties about environmental outcomes (see, for example, Brookshire et al. 1983).

Furthermore, in some cases, the explicit goal of a policy or other action might be a reduction in a specific health or environmental risk.[19] For example, a reduction in air pollution can reduce the risk of contracting an illness such as asthma. It can also reduce mortality rates, particularly for infants (e.g., Currie and Neidell 2005; Agarwal et al. 2010). Thus, the benefits of reductions in air pollution include reductions in these health risks. Values must be placed on the risk reductions to estimate these benefits. As with other types of benefits, the basic concept of economic value can be applied to the value of these risk reductions as well, for example, by estimating the amount that individuals would be willing to pay to reduce these risks (e.g., Cameron and DeShazo 2013).[20]

Therefore, when the risks or uncertainties can be quantified and are borne by the individuals assigning values, the standard approach to nonmarket valuation can be modified to explicitly incorporate uncertainty. To date, however, efforts to incorporate uncertainty into nonmarket valuation have been primarily limited to the valuation of health-related risks. The valuation of other types of environmental risks, particularly those that are long term, geographically broad, and potentially catastrophic, remains a significant challenge.

1.9 Valuation Methods

As noted, this book is primarily about nonmarket valuation methods that can be used in Step 5 of the valuation process depicted in Table 1.1. A number of nonmarket valuation methods exist. Most of them have a long history of use within the field of environmental and natural resource economics, while others (such as the experimental methods discussed in Chap. 10) are newer. All of them seek to estimate the economic values individuals assign to goods and services that are not traded in markets (such that values cannot be directly inferred from market prices).

Although all of the methods described in this book seek to estimate economic values, they differ in a number of ways, including the following:

[19] Although some authors distinguish between "risk" and "uncertainty" based on whether the probabilities of the possible outcomes can be quantified, in economics, the two terms are typically used interchangeably. We follow this convention here. For example, when talking about either the uncertainty associated with future preferences or the health risks from exposure to pollution, we assume that all possibilities can be identified and each can be assigned an objective (or possibly subjective) probability of occurring.

[20] Such estimates are often expressed in terms of the "value of a statistical life" (VSL). However, this terminology has led to confusion and unnecessary controversy about the concept being measured, prompting some to argue that the term VSL should not be used (e.g., Cameron 2010).

Table 1.2 Major nonmarket valuation methods

Revealed preference	Stated preference
Travel cost	Contingent valuation
Hedonics	Attribute-based methods
Defensive behavior	
Substitution methods	

1. Revealed preference methods estimate values by observing actual behavior that is linked in some way to an environmental good or attribute (such as visits to a recreational site or the purchase of a home), and then inferring values indirectly from that behavior. Stated preference methods estimate values by asking individuals survey questions related to their preferences and inferring values from their stated responses. As a result, the types of data used differ across methods. Revealed preference methods rely on observed data, which may include data collected through surveys related to behavior or market outcomes (for example, data on visits to sites or house prices). In contrast, stated preference methods require surveys that use hypothetical questions designed specifically to elicit information about values (for example, questions about willingness to pay or questions that require a choice among hypothetical alternatives). Table 1.2 categorizes the major nonmarket valuation methods based on this distinction.
2. Methods also differ in terms of the components of total economic value they can capture. For example, revealed preference methods can only capture use values, while stated preference methods can (in principle) estimate both use and passive-use values. Likewise, specific revealed preference methods capture only certain kinds of use value. For example, hedonic methods capture only use values that are capitalized into prices of related goods or services (such as housing), while travel cost methods capture only the value of goods that require travel to a site.
3. The resources (both time and money) needed to do the valuation can also differ, depending on the method chosen. This implies that some methods might be more feasible or appropriate (depending on information needs) in some contexts than others. For example, major regulations might warrant the significant expenditures involved in doing an original valuation-related survey, while minor regulations or less significant decisions might be able to rely on less resource-intensive benefit transfers.
4. Finally, in some contexts, methods might differ in terms of their perceived acceptability as a reliable means of estimating values. For example, regulatory impact analyses might need to rely primarily on methods that have been deemed acceptable or are explicitly preferred by those with authority/responsibility for regulatory review.[21] Similarly, there continues to be debate about acceptability of

[21]For example, the U.S. Office of Management and Budget's Circular A-4, which governs federal regulatory impact analyses, explicitly states: "Other things equal, you should prefer revealed preference data over stated preference data because revealed preference data are based on actual

the use of stated preference methods in the estimation of natural resource damage assessments and benefit-cost analyses (e.g., Hausman 2012; Kling et al. 2012).

These differences across the available methods imply that no single method will be suitable for all valuation needs. Rather, the choice of method must be context-specific. In addition, it may be necessary or desirable to use multiple methods in some contexts. First, combining information from stated and revealed preference methods can sometimes improve benefit estimation of a single component (see, for example, Adamowicz et al. 1994). Second, because revealed preference methods do not provide complete estimates of total economic value, different components of total economic value sometimes can be estimated using different methods and then aggregated, although care must be taken to avoid double counting if the components of value captured by the different methods overlap.

1.10 Outline of the Book

This book is designed to provide a basic understanding of nonmarket valuation. Chapter 2 presents an overview of the economic theory that underlies all of the nonmarket valuation methods discussed here. It is important to understand conceptually what nonmarket valuation methods seek to estimate to ensure that the methods are applied correctly and the resulting estimates are appropriately understood, interpreted, and used.

Although Chap. 2 is theoretical, the emphasis of the book is on the use and application of nonmarket valuation. Data are a critical part of any application, and Chap. 3 provides a discussion about collecting survey data for use in nonmarket valuation. The issues discussed in this chapter are relevant to a variety of valuation methods and serve as additional background (along with the theory in Chap. 2) for the discussion of specific methods in the chapters that follow. These include both stated preference methods (Chaps. 4 and 5) and revealed preference methods (Chaps. 6 through 9). For each method, both fundamentals and recent advances are discussed.

While Chaps. 4 through 9 focus on specific methods, Chaps. 10 and 11 address more broadly the use of experiments as part of a valuation exercise (Chap. 10) and the application of benefit transfer as an alternative to doing an original valuation study (Chap. 11). The use of experiments is a rapidly growing field within economics, and the intersection between experimental methods and nonmarket valuation is a very promising, emerging area. Similarly, although benefit transfer is a well-established approach and has been used extensively, its importance is likely to

(Footnote 21 continued)

decisions, where market participants enjoy or suffer the consequences of their decisions". See OMB Circular A-4, Section E (September 17, 2003), available at http://www.whitehouse.gov/omb/circulars_a004_a-4#e.

increase in the future as decision-makers increasingly seek information about the value of environmental goods and services in contexts where conducting an original valuation study is not feasible or warranted.

Finally, although the methods discussed in this book are firmly grounded in economic theory and have a long history of use, as with any estimation method, assessing the validity of the resulting estimates is an important part of ensuring that they are appropriately understood, interpreted, and used. Chapter 12 discusses general issues related to assessing the validity of value estimates derived from nonmarket valuation. Although no method is perfect, validity checks can increase the confidence users have in the value estimates that emerge. This should, in turn, increase their confidence that use of those estimates will lead to better decisions—the ultimate goal of nonmarket valuation.

References

Adamowicz, W. (2004). What's it worth? An examination of historical trends and future directions in environmental valuation. Australian Journal of Agricultural and Resource Economics, 48, 419-443.

Adamowicz, W., Louviere, J. & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. Journal of Environmental Economics and Management, 26, 271-292.

Agarwal, N., Banternghansa, C. & Bui, L. T. M. (2010). Toxic exposure in America: Estimating fetal and infant health outcomes from 14 years of TRI reporting. Journal of Health Economics, 29, 557-574.

Arrow, K. J., Cropper, M. L., Eads, G. C., Hahn, R. W., Lave, L. B., Noll, R. G., et al. (1996). Is there a role for benefit-cost analysis in environmental, health, and safety regulation? Science, 272 (April), 221-222.

Benhabib, J., Bisin, A. & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. Games and Economic Behavior, 69, 205-223.

Bockstael, N. E., Freeman, A. M. III, Kopp, R. J., Portney, P. R. & Smith, V. K. (2000). On measuring economic values for nature. Environmental Science and Technology, 34, 1384-1389.

Brookshire, D. S., Eubanks, L. S. & Randall, A. (1983). Estimating option prices and existence values for wildlife resources. Land Economics, 59, 1-15.

Brown, T. C. (1984). The concept of value in resource allocation. Land Economics, 60, 231-246.

Brown, T. C., Bergstrom, J. C. & Loomis, J. B. (2007). Defining, valuing, and providing ecosystem goods and services. Natural Resources Journal, 47, 329-376.

Cameron, T. A. (2010). Euthanizing the value of a statistical life. Review of Environmental Economics and Policy, 4, 161-178.

Cameron, T. A. & DeShazo, J. R. (2013). Demand for health risk reductions. Journal of Environmental Economics and Management, 65, 87-109.

Carpenter, S., Defries, R., Dietz, T., Mooney, H. A., Polasky, S., Reid, W. V., et al. (2006). Millennium ecosystem assessment: Research needs. Science, 314, 257-258.

Carson, R. T. (2012). Contingent valuation: A comprehensive bibliography and history. Cheltenham, United Kingdom: Edward Elgar Publishing.

Cerda, C., Barkmann, J. & Marggraf, R. (2013). Application of choice experiments to quantify existence value of an endemic moss: A case study in Chile. Environment and Development Economics, 18, 207-224.

Common, M., Reid, I. & Blamey, R. (1997). Do existence values for cost benefit analysis exist? Environmental and Resource Economics, 9, 225-238.
Cooper, P., Poe, G. L. & Bateman, I. J. (2004). The structure of motivation for contingent values: A case study of lake water quality improvement. Ecological Economics, 50, 69-82.
Costanza, R., d'Arge, R., de Groot, R., Farber, S. & Grasso, M., Hannon, et al. (1997). The value of the world's ecosystem services and natural capital. Nature, 387, 253-260.
Cropper, M. L. (2012). How should benefits and costs be discounted in an intergenerational context? Resource for the Future Discussion Paper, RFF DP 12-42, Washington, DC.
Currie, J. & Neidell, M. (2005). Air pollution and infant health: What can we learn from California's recent experience? The Quarterly Journal of Economics, 120, 1003-1030.
Daily, G. C. (Ed.) (1997). Nature's services: Societal dependence on natural ecosystems. Washington, DC: Island Press.
Daily, G. & Ellison, K. (2002). The new economy of nature: The quest to make conservation profitable. Washington, DC: Island Press.
Dietz, T., Fitzgerald, A. & Shwom, R. (2005). Environmental values. Annual Review of Environment and Resources, 30, 335-372.
Freeman, A. M. III., Herriges, J.A. & Kling, C.L. (2014). The measurement of environmental and resource values: Theory and methods (3rd ed.). Washington, DC: RFF Press.
Hausman, J. (2012). Contingent valuation: From dubious to hopeless. Journal of Economic Perspectives, 26 (4), 43-56.
Heal, G. (2000). Nature and the marketplace: Capturing the value of ecosystem services. Washington, DC: Island Press.
Heal, G. (2005). Intertemporal welfare economics and the environment. In K.-G. Mäler & J. Vincent (Eds.), Handbook of environmental economics, Vol. 3: Economywide and international environmental issues (pp. 1105-1145). Amsterdam: Elsevier.
Heal, G. (2009). Climate economics: A meta-review and some suggestions for future research. Review of Environmental Economics and Policy, 3, 4-21.
Johansson-Stenman, O. (2005). Distributional weights in cost-benefit analysis: Should we forget about them? Land Economics, 81, 337-352.
Johnston, R. J., Segerson, K., Schultz, E. T., Besedin, E. Y. & Ramachandran, M. (2011). Indices of biotic integrity in stated preference valuation of aquatic ecosystem services. Ecological Economics, 70, 1946-1956.
Just, R. E., Hueth, D. L. & Schmitz, A. (2004). The welfare economics of public policy: A practical approach to project and policy evaluation. Cheltenham, United Kingdom: Edward Elgar Publishing.
Kling, C. L., Phaneuf, D. J. & Zhao, J. (2012). From Exxon to BP: Has some number become better than no number? Journal of Economic Perspectives, 26 (4), 3-26.
Kopp, R. J. & Smith, V. K. (Eds.) (1993). Valuing natural assets: The economics of natural resource damage assessment. Washington, DC: RFF Press.
Lichtenstein, S. & Slovic, P. (Eds.) (2006). The construction of preferences. New York: Cambridge University Press.
McVittie, A. & Moran, D. (2010). Valuing the non-use benefits of marine conservation zones: An application to the UK Marine Bill. Ecological Economics, 70, 413-424.
Millennium Ecosystem Assessment. (2005). Ecosystems and human well-being: A framework for assessment. Washington, DC: Island Press.
National Research Council. (2005). Valuing ecosystem services: Toward better environmental decision-making. Washington, DC: National Academies Press.
Polasky, S. & Segerson, K. (2009). Integrating ecology and economics in the study of ecosystem services: Some lessons learned. Annual Review of Resource Economics, 1, 409-434.
Portney, P. R. (1994). The contingent valuation debate: Why economists should care. Journal of Economic Perspectives, 8 (4), 3-17.
Smith, V. K. (Ed.) (1984). Environmental policy under Reagan's executive order: The role of benefit-cost analysis. Chapel Hill: University of North Carolina Press.

Smith, V. K. (1993). Nonmarket valuation of environmental resources: An interpretive appraisal. Land Economics, 69, 1-26.

Spash, C. L. (2006). Non-economic motivation for contingent values: Rights and attitudinal beliefs in the willingness to pay for environmental improvements. Land Economics, 82, 602-622.

Toman, M. (1998). Why not to calculate the value of the world's ecosystem services and natural capital. Ecological Economics, 25, 57-60.

U.S. Environmental Protection Agency. (2009). Valuing the protection of ecological systems and services: A report of the EPA Science Advisory Board. EPA-SAB-09-012, Washington, DC.

Viscusi, W. K. (2009). Valuing risks of death from terrorism and natural disasters. Journal of Risk and Uncertainty, 38, 191-213.

Zhao, J. & Kling, C. L. (2009). Welfare measures when agents can learn: A unifying theory. The Economic Journal, 119, 1560-1585.

Chapter 2
Conceptual Framework for Nonmarket Valuation

Nicholas E. Flores

Abstract This chapter provides an overview of the theoretical foundations of nonmarket valuation. The chapter first develops a model of individual choice where private goods are freely chosen but environmental goods are rationed from the individual's perspective. The model is used to define compensating and equivalent welfare measures for changes in prices and environmental goods. These welfare measures form the basis of the environmental values researchers seek to measure through nonmarket valuation. The chapter discusses the travel cost model with and without weak complementarity, the household production model, the hedonic model, and the general concept of passive-use value. The individual choice model is extended to a dynamic framework and separately to choice under uncertainty. Finally the chapter develops welfare measures associated with averting expenditures and random utility models.

Keywords Public goods · Welfare economics · Compensating welfare measures · Equivalent welfare measures · Weak complementarity · Passive-use value · Uncertainty · Averting expenditures · Random utility model

Serious practice of nonmarket valuation requires a working knowledge of the underlying economic theory because it forms the basis for the explicit goals in any nonmarket valuation exercise. This chapter provides readers with the requisite theory to meaningfully apply the nonmarket valuation techniques described in this book.

To do so, this chapter develops a model of individual choice that explicitly recognizes the public good nature of many applications. While the emphasis is on public goods, the concepts in this chapter and the methods in this book have broader applicability to newly introduced market goods and goods that are not pure public goods. This model is used to derive the basic welfare measures that nonmarket valuation studies measure. Moving toward a more specific framework, the

N.E. Flores (✉)
University of Colorado-Boulder, Boulder, USA
e-mail: Nicholas.Flores@Colorado.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_2

chapter examines how market behavior can be used to identify the basic welfare measures for nonmarket goods. It also provides a discussion of situations for which market demands are not sufficient to recover the basic welfare measures, cases of passive-use value, and visits to new recreation sites. That is followed by a discussion of intertemporal choice and nonmarket valuation, nonmarket valuation under uncertainty, use of averting expenditures to value nonmarket goods, and, finally, welfare measures for discrete-choice, random utility models.[1]

2.1 Theoretical Model of Nonmarket Goods

The chapter begins with some illustrative examples. Air quality, water quality of lakes and streams, and the preservation of public lands are relevant examples of nonmarket goods. Each of these goods can change due to society's choices, but individuals may not unilaterally choose their preferred level of air quality, water quality, or acreage of preserved public lands. In addition to being outside of the choice set of any individual, these examples have the common feature that everyone experiences the same level of the good. Citizens at a given location experience the same level of local air quality; citizens of a state or province experience the same level of water quality in the state's lakes and streams; and everyone shares the level of preserved public lands. People can choose where to live or recreate, but environmental quality at specific locations is effectively rationed. Rationed, common-level goods serve as the point of departure for standard neoclassical price theory in developing the theoretical framework for nonmarket valuation.

The basic premise of neoclassical economic theory is that people have preferences over goods—in this case, both market and nonmarket goods. Without regard to the costs, each individual is assumed to be able to order bundles of goods in terms of desirability, resulting in a complete preference ordering. The fact that each individual can preference order the bundles of goods forms the basis of choice. The most fundamental element of economic theory is the preference ordering, or more simply, the desires of the individual—not money. Money plays an important role because individuals have a limited supply of money to buy some, but not all, of the things they want. An individual may desire improved air or water quality or the preservation of an endangered species for any reason, including personal use, bequests to future generations, or simply for the existence of the resource. Economic theory is silent with regard to motivation. As Becker (1993, p. 386) offered, the reasons for enjoyment of any good can be "selfish, altruistic, loyal, spiteful, or masochistic." Economic theory provides nearly complete flexibility for accommodating competing systems of preferences.

[1]These topics alone could constitute an entire book, but the treatment of each must be brief. For those launching a career in this area, Freeman (1993) and Hanley et al. (1997) are recommended.

Preference ordering can be represented through a utility function defined over goods. For these purposes, $X = [x_1, x_2, \cdots, x_n]$ denotes a list or vector of all of the levels for the n market goods the individual chooses. The k nonmarket goods are similarly listed as $Q = [q_1, q_2, \ldots, q_k]$. The utility function assigns a single number, $U(X, Q)$, for each bundle of goods (X, Q). For any two bundles (X^A, Q^A) and (X^B, Q^B), the respective numbers assigned by the utility function are such that $U(X^A, Q^A) > U(X^B, Q^B)$ if and only if (X^A, Q^A) is preferred over (X^B, Q^B). The utility function is thus a complete representation of preferences.[2]

Money enters the process through scarcity and, in particular, scarcity of money to spend on obtaining the things we enjoy, i.e., a limited budget. For market goods, individuals choose the amount of each good to buy based on preferences, the relative prices of the market goods $P = (p_1, p_2, \ldots, p_n)$, and available income. Given this departure point, the nonmarket goods are rationed in the sense that individuals may not unilaterally choose the level of these goods.[3] The basic choice problem is how to obtain the highest possible utility level when spending income y toward the purchase of market goods is subject to a rationed level of the nonmarket goods:

$$\max_X U(X, Q) \text{ s.t. } P \cdot X \le y, Q = Q^0. \tag{2.1}$$

There are two constraints that people face in Eq. (2.1). First, the total expenditure on market goods cannot exceed income (budget constraint),[4] and second, the levels of the nonmarket goods are fixed.[5] The X that solves this problem then depends on the level of income (y), the prices of all of the market goods (P), and the level of the rationed, nonmarket goods(Q). For each market good, there is an optimal demand function that depends on these three elements, $x_i^* = x_i(P, Q, y)$. The vector of optimal demands can be written similarly, $X^* = X(P, Q, y)$, where the vector now lists the demand function for each market good. If one plugs the set of optimal demands into the utility function, he or she obtains the indirect utility function $U(X^*, Q) = v(P, Q, y)$. Because the demands depend on prices, the levels of the nonmarket goods, and income, the highest obtainable level of utility also depends on these elements.

As the name suggests, demand functions provide the quantity of goods demanded at a given price vector and income level. Demand functions also can be

---

[2]The utility function is ordinal in the sense that many different functions could be used to equally represent a given preference ordering. For a complete discussion of preference orderings and their representations by utility functions, see Kreps (1990) or Varian (1992).

[3]One can choose goods that have environmental quality attributes, e.g., air quality and noise. These goods are rationed in the sense that an individual cannot unilaterally improve ambient air quality or noise level at his or her current house. One can move to a new location where air quality is better but cannot determine the level of air quality at his or her current location.

[4]It may be the case that one has to pay for Q^0. Rather than including this payment in the budget constraint, he or she can simply consider income to already be adjusted by this amount. Because the levels of the nonmarket goods are not individually chosen, there is no need to include payments for nonmarket goods in the budget constraint.

[5]To clarify notation, $p \cdot X = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$, where p_i is the price of market good i.

interpreted as marginal value curves because consumption of goods occurs up to the point where marginal benefits equal marginal costs. For this reason, demand has social significance.

2.1.1 Compensating and Equivalent Welfare Measures

Policies or projects that provide nonmarket goods often involve costs. Values may be assigned to these policies or projects in order to assess whether the benefits justify the costs. For example, consider a policy intended to improve the water quality of Boulder Creek, a stream that runs through my hometown of Boulder, Colo. I care about this stream because I jog along its banks and enjoy the wildlife it supports, including the trout my daughters may catch when they are lucky. To pay for a cleanup of this creek, the prices of market goods might change due to an increase in sales tax, and/or I might be asked to pay a lump sum fee.

Two basic measures of value that are standard fare in welfare economics can be used to assess the benefit of cleaning up Boulder Creek. The first is the amount of income I would give up after the policy has been implemented that would exactly return my utility to the status quo utility level before cleanup. This measure is the "compensating" welfare measure, which is referred to as C. Letting "0" superscripts denote the initial, status quo conditions and "1" superscripts denote the new conditions provided by the policy, C is generally defined using the indirect utility function as follows:

$$v(P^0, Q^0, y^0) = v(P^1, Q^1, y^1 - C). \tag{2.2}$$

The basic idea behind C is that if I give up C at the same time I experience the changes $(P^0, Q^0, y^0) \rightarrow (P^1, Q^1, y^1)$, then I am back to my original utility. My notation here reflects a general set of changes in prices, rationed nonmarket goods, and income. In many cases, including the example of water quality in Boulder Creek, only environmental quality is changing. C could be positive or negative, depending on how much prices increase and/or the size of any lump sum tax I pay. If costs are less than C and the policy is implemented, then I am better off than before the policy. If costs are more than C, I am worse off.

The second basic welfare measure is the amount of additional income I would need with the initial conditions to obtain the same utility as after the change. This is the *equivalent* welfare measure, referred to as E, and is defined as

$$v(P^0, Q^0, y^0 + E) = v(P^1, Q^1, y^1). \tag{2.3}$$

The two measures differ by the implied assignment of property rights. For the compensating measure, the initial utility level is recognized as the basis of comparison. For the equivalent measure, the subsequent level of utility is recognized as the basis. Whether one should consider the compensating welfare measure or the equivalent welfare measure as the appropriate measure depends on the situation.

Suppose a new policy intended to improve Boulder Creek's water quality is being considered. In this case, the legal property right is the status quo; therefore, the analyst should use the compensating welfare measure. There are, however, instances when the equivalent welfare measure is conceptually correct. Returning to the water quality example, in the U.S., the Clean Water Act provides minimum water quality standards. If water quality declined below a standard and the project under consideration would restore quality to this minimum standard, then the equivalent welfare measure is the appropriate measure. Both conceptual and practical matters should guide the choice between the compensating and equivalent welfare measure.[6]

2.1.2 Duality and the Expenditure Function

So far, the indirect utility function has been used to describe the basic welfare measures used in economic policy analysis. To more easily discuss and analyze specific changes, the analyst can equivalently use the expenditure function to develop welfare measures. The indirect utility function represents the highest level of utility obtainable when facing prices P, nonmarket goods Q, and income y.

Expenditure minimization is the flip side of utility maximization and is necessary for utility maximization. To illustrate this, suppose an individual makes market good purchases facing prices P and nonmarket goods Q and obtains a utility level of U^0. Now suppose he or she is not minimizing expenditures, and U^0 could be obtained for less money through a different choice of market goods. If this were true, the person would not be maximizing utility because he or she could purchase the alternative, cheaper bundle that provides U^0 and use the remaining money to buy more market goods and, thus, obtain a utility level higher than U^0. This reasoning is the basis of what microeconomics refers to as "duality." Instead of looking at maximizing utility subject to the budget constraint, the dual objective of minimizing expenditures—subject to obtaining a given level of utility—can be considered. The expenditure minimization problem is stated as follows:

$$\min_X P \cdot X \, s.t. \, U(X, Q) \geq U^0, Q = Q^0. \tag{2.4}$$

The solution to this problem is the set of compensated or Hicksian demands that are a function of prices, nonmarket goods levels, and level of utility,

[6]Interest over the difference in size between C and E has received considerable attention. For price changes, Willig (1976) provided an analysis. For quantity changes, see Randall and Stoll (1980) and Hanemann (1991). Hanemann (1999) provided a comprehensive and technical review of these issues. From the perspective of measurement, there is a general consensus that it is more difficult to measure E, particularly in stated preference analysis.

$X^* = X^h(P, Q, U)$. The dual relationship between the ordinary demands and the Hicksian demands is that they intersect at an optimal allocation $X(P, Q, y) = X^h(P, Q, U)$ when $U = v(P, Q, y)$ in the expenditure minimization problem and $y = P \cdot X^h(P, Q, y)$ in the utility maximization problem.

As the term "duality" suggests, these relationships represent two views of the same choice process. The important conceptual feature of the compensated demands is that utility is fixed at some specified level of utility, which relates directly to our compensating and equivalent welfare measures. For the expenditure minimization problem, the expenditure function, $e(P, Q, y) = P \cdot X^h(P, Q, U)$, takes the place of the indirect utility function.

It is worth stressing that the expenditure function is the ticket to understanding welfare economics. Not only does the conceptual framework exactly match the utility-constant nature of welfare economics, the expenditure function itself has very convenient properties. In particular, the expenditure function approach allows one to decompose a policy that changes multiple goods or prices into a sequence of changes that will be shown to provide powerful insight into our welfare measures.

This chapter has so far introduced the broad concepts of compensating and equivalent welfare measures. Hicks (Hicks 1943) developed the compensating and equivalent measures distinctly for price and quantity changes and named them the price compensating/equivalent variation for changes in prices and the quantity compensating/equivalent variation for quantity changes, respectively. These two distinct measures are now typically referred to as the compensating/equivalent variation for price changes and the compensating/equivalent surplus for quantity changes. It is easy to develop these measures using the expenditure function, particularly when one understands the terms "equivalent" and "compensating."

Before jumping directly into the compensating/equivalent variations and surpluses, income changes should be discussed. Income changes can also occur as a result of policies, so changes in income are discussed first. For example, regulating the actions of polluting firms may decrease the demand for labor and result in lower incomes for workers.

2.1.3 The Treatment of Income Changes

Let $U^0 = v(P^0, Q^0, y^0)$ represent the status quo utility level and $U^1 = v(P^1, Q^1, y^1)$ the utility level after a generic change in income, prices, and/or nonmarket goods. The two measures are defined by the fundamental identities as follows:

$$v(P^0, Q^0, y^0) = v(P^1, Q^1, y^1 - C) \tag{2.5a}$$

$$v(P^0, Q^0, y^0 + E) = v(P^1, Q^1, y^1) \tag{2.5b}$$

Also, C and E can be represented using the expenditure function:

$$C = e(P^1, Q^1, U^1) - e(P^1, Q^1, U^0), \tag{2.6a}$$

$$E = e(P^0, Q^0, U^1) - e(P^0, Q^0, U^0) \tag{2.6b}$$

To determine how to handle income changes, C and E need to be rewritten in more workable forms. In expenditure terms, $y^0 = e(P^0, Q^0, U^0)$, $y^1 = e(P^1, Q^1, U^1)$, and $y^1 = y^0 + y^1 - y^0$. By creatively using these identities, C and E can be rewritten as

$$C = e\left(P^0, Q^0, U^0\right) - e\left(P^1, Q^1, U^0\right) + (y^1 - y^0). \tag{2.7a}$$

$$E = e\left(P^0, Q^0, U^1\right) - e\left(P^1, Q^1, U^1\right) + (y^1 - y^0) \tag{2.7b}$$

The new form shows that for C, one values the changes in prices and nonmarket goods at the initial utility level and then considers the income change. For E, one values the changes in prices and nonmarket goods at the post-change utility level and then considers income change. The generalized compensated measure is subtracted from income under the subsequent conditions (Eq. 2.2), while the generalized equivalent measure is added to income under the initial conditions (Eq. 2.3), regardless of the direction of changes in P or Q. How the changes in prices and nonmarket goods are valued is the next question.

2.1.4 *Variation Welfare Measures for a Change in Price* i

Suppose the analyst is considering a policy that only provides a price increase for good i. Hicks (1943) referred to the compensating welfare measure for a price change as "compensating variation" (CV) and to the equivalent welfare measure as "equivalent variation" (EV). Because a price decrease makes the consumer better off, both measures are positive. P_{-i} refers to the price vector left after removing p_i:

$$CV = e\left(p_i^0, P_{-i}^0, Q^0, U^0\right) - e\left(p_i^1, P_{-i}^0, Q^0, U^0\right); \tag{2.8}$$

$$EV = e\left(p_i^0, P_{-i}^0, Q^0, U^1\right) - e\left(p_i^1, P_{-i}^0, Q^0, U^1\right). \tag{2.9}$$

Using Roy's identity and the fundamental theorem of calculus, compensating and equivalent variations can be expressed as the area under the Hicksian demand curve between the initial and subsequent price.[7] Here, s represents p_i along the path of integration:

[7]Roy's identity states that the derivative of the expenditure function with respect to price i is simply the Hicksi demand for good i. The fundamental theorem of calculus allows one to write the difference of two differentiable functions as the integral over the derivative of that function.

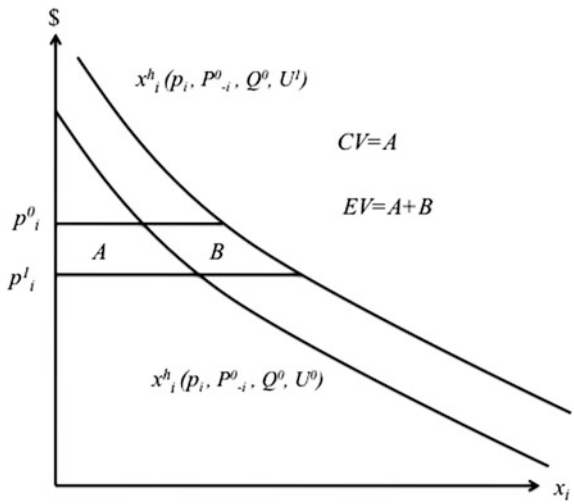


Fig. 2.1 Compensating and equivalent variations for a decrease in p_i

$$CV = e\left(p_i^0, P_{-i}^0, Q^0, U^0\right) - e\left(p_i^1, P_{-i}^0, Q^0, U^0\right) = \int_{p_i^1}^{p_i^0} x_i^h\left(s, P_{-i}^0, U^0\right) ds; \tag{2.10}$$

$$EV = e\left(p_i^0, P_{-i}^0, Q^0, U^1\right) - e\left(p_i^1, P_{-i}^0, Q^0, U^1\right). = \int_{p_i^1}^{p_i^0} x_i^h\left(s, P_{-i}^0, U^1\right) ds. \tag{2.11}$$

For the price change, compensating variation is simply the area under the Hicksian demand curve evaluated at the initial utility level and the two prices. Similarly, equivalent variation is simply the area under the Hicksian demand curve evaluated at the new utility level and the two prices. Figure 2.1 depicts these two measures for the price change.

A few issues regarding the welfare analysis of price changes deserve mention. First, only a single price change has been presented. Multiple price changes are easily handled using a compensated framework that simply decomposes a multiple price change into a sequence of single price changes (Braeutigam and Noll 1984). An example of how to do this is provided in the discussion of weak in Sect. 2.2.2. Second, the area under the ordinary (uncompensated) demand curve and between the prices is often used as a proxy for either compensating or equivalent variation. Willig (1976) had shown that in many cases this approximation is quite good, depending on the income elasticity of demand and the size of the price change. Hausman (1981) offered one approach to deriving the exact Hicksian measures from ordinary demands. Vartia (1983) offered another approach that uses numerical methods for deriving the exact Hicksian measures. While both methods for deriving the compensated welfare measures from ordinary demands are satisfactory, Vartia's method is very simple.

Finally, the analyst also needs to consider price increases, which are conceptually the same except that the status quo price is now the lower price, $P^0 < P^1$. Both welfare measures here are negative. In the case of compensating variation, an individual takes away a negative amount, i.e., gives money, because the new price level makes him or her worse off. Similarly, one would have to give up money at the old price in order to equate the status quo utility with the utility at the new price, which is equivalent to saying a negative equivalent variation exists.

2.1.5 Welfare Measures for a Change in Nonmarket Goods

Now suppose one is considering an increase in the amount of the nonmarket good $q_{j.}$ This change could represent acres of open space preserved, something that most would consider a quantity change, or the level of dissolved oxygen in a stream, a quality change that can be measured. Recall that the compensating and equivalent measures are referred to as compensating surplus (CS) and equivalent surplus (ES). The expenditure function representation of these is given as follows:

$$CS = e(P^0, Q^0, U^0) - e(P^0, Q^1, U^0); \tag{2.12}$$

$$ES = e(P^0, Q^0, U^1) - e(P^0, Q^1, U^1). \tag{2.13}$$

Using the properties of the expenditure function, one can rewrite the quantity compensating and equivalent variations in an insightful form. Maler (1974) showed that the derivative of the expenditure function with respect to nonmarket good q_j is simply the negative of the inverse Hicksian demand curve for nonmarket good q_j. This derivative equals the negative of the virtual price—the shadow value—of nonmarket good q_j. Again applying the fundamental theorem of calculus, the analyst can rewrite the surplus measures in terms of this shadow value. Similar to the notation for price changes, Q_{-j} refers to the price vector left after removing q_j, and s represents q_j along the path of integration.

$$\begin{aligned} CS &= e(P^0, q_j^0, Q_{-j}^0, U^0) - e(P^0, q_j^1, Q_{-j}^0, U^0) \\ &= \int_{q_j^0}^{q_j^1} p_i^v\left(P^0, s, Q_{-j}^0, U^0\right) ds; \end{aligned} \tag{2.14}$$

$$\begin{aligned} ES &= e\left(P^0, q_j^0, Q_{-j}^0, U^1\right) - e\left(P^0, q_j^1, Q_{-j}^0, U^1\right) \\ &= \int_{q_j^0}^{q_j^1} p_i^v\left(P^0, s, Q_{-j}^0, U^1\right) ds. \end{aligned} \tag{2.15}$$

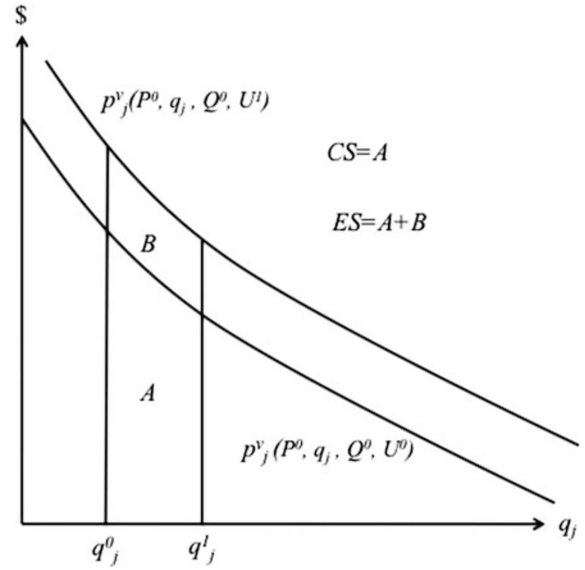


Fig. 2.2 Compensating and equivalent surpluses for an increase in q_j

Figure 2.2 graphs the compensating and equivalent surpluses for this increase in nonmarket good q_j. The graph looks similar to Fig. 2.1 except that the change is occurring in the quantity space as opposed to the price space. For normal nonmarket goods—goods where the quantity desired increases with income—the equivalent measure will exceed the compensating measure for increases in the nonmarket good. For decreases in the nonmarket good, the opposite is true.

In thinking about compensating/equivalent surpluses as opposed to the variations, it is useful to remember what is public and what is private. In the case of market goods, prices are public, and the demand for the goods varies among individuals. For nonmarket goods, the levels are public and shared by all, while the marginal values vary among individuals. These rules of thumb help to differentiate between the graphic representations of compensating and equivalent variations and surpluses.

Table 1 Willingness to pay and willingness to accept

Welfare measure	Price increase	Price decrease
Equivalent variation—Implied property right in the change	WTP to avoid	WTA to forgo
Compensating variation—Implied property right in the status quo	WTA to accept	WTP to obtain

Source Freeman [1993, p. 58)

2.1.6 Compensating and Equivalent Variations and Willingness to Pay and Willingness

Two other terms, "willingness to pay" (WTP) and "willingness to accept" (WTA) compensation, are often used as substitute names for either the compensating measures or the equivalent measures. WTP is typically associated with a desirable change, and WTA compensation is associated with a negative change. Consider Table 2.1 for a price change.

As the table suggests, one needs to be explicit about what he or she is paying for when using WTP, and one needs to be explicit about what he or she is being compensated for when using WTA. In cases where utility changes are unambiguously positive or negative, the WTP/WTA terminology works well. However, when combinations of desirable and undesirable changes exist, such as an increase in water quality accompanied by an increase in sales taxes on market goods, then WTP and WTA are less useful terms. This is true because if the policy as a whole is bad $(U^0 > U^1)$, then the compensating welfare measure is WTA, and the equivalent welfare measure is WTP to avoid the policy. If the policy as a whole is good $(U^0 < U^1)$, then the compensating welfare measure is WTP to obtain the policy, and the equivalent welfare measure is WTA to forgo the policy. The situation could result in mixed losses and gains, leading one to measure WTA for losers and WTP for gainers, using the WTP/WTA terminology. Using equivalent or compensating welfare measures, one measure is used for losers and gainers. Hanemann (1991, 1999) provided theoretical and empirical evidence that the difference between compensating and equivalent measures can be quite dramatic. WTP for the increase in a unique nonmarket good that has virtually no substitutes can be many orders of magnitude smaller than WTA compensation to give up the increase.

These concepts refer to gains and losses at the individual level. There are different approaches to aggregating information from individuals to make collective choices. The Kaldor–Hicks criterion is the most widely used approach to aggregating compensating or equivalent welfare measures. A proposed change passes the Kaldor test if the sum of the compensating measures is greater than zero; the proposed change passes the Hicks test if the sum of equivalent measures is greater than zero.

As noted by Freeman (1993), the choice of test depends on the decision context. The Kaldor–Hicks criterion implies that projects passing the selected test satisfy the requirement that the gains of winners are more than sufficient to compensate the losers, leading to the potential that the change could occur with redistribution of income where some gain and none lose. It is important to recognize that compensation need not occur.

2.2 Implicit Markets for Environmental Goods

By definition, individuals do not explicitly purchase nonmarket goods. They do, however, purchase other goods for which demands are related to nonmarket goods. For example, one's choice of where to recreate may depend on the environmental quality of the sites under consideration. Furthermore, environmental quality can influence one's choice of which community to live in or which house to buy once he or she has decided on a community. These market links to nonmarket goods make it possible to infer values for the demand revealed through these purchases. The specific nonmarket valuation techniques used to infer these values, called revealed preference methods, are described in Chaps. 6 through 8. Section 2.2 reviews some of the concepts related to inferring environmental values from market purchases.

2.2.1 Price Changes and Environmental Values

This section develops a framework that relates changes in nonmarket goods to price changes in market goods. This is done in order to introduce the weak complementarity condition, a condition that, if satisfied, allows changes in nonmarket goods to be valued through changes in consumer surplus of affected market goods. Suppose one is increasing the first nonmarket good q_1, wishes to measure the monetary value for this change, and determines compensating surplus to be the appropriate measure. Using the expenditure function, the only argument that changes is q_1. Q_{-1} is the vector left after removing the first element of Q:

$$CS = e\left(P^0, q_1^0, Q_{-1}^0, U^0\right) - e\left(P^0, q_1^1, Q_{-1}^0, U^0\right). \tag{2.16}$$

The next step is the introduction of an arbitrary price change along with this quantity change by adding and subtracting two different terms. The size of the compensating surplus has not changed:

$$\begin{aligned} CS = {} & e\left(P^1, q_1^1, Q_{-1}^0, U^0\right) - e\left(P^0, q_1^1, Q_{-1}^0, U^0\right) \\ & - \left[e\left(P^1, q_1^0, Q_{-1}^0, U^0\right) - e\left(P^0, q_1^0, Q_{-1}^0, U^0\right)\right] \\ & + e\left(P^1, q_1^0, Q_{-1}^0, U_0\right) - e\left(P^1, q_1^1, Q_{-1}^0, U^0\right). \end{aligned} \tag{2.17}$$

The second and fourth terms are the original terms in (2.16) and the other four are the "zero" terms. Note the arrangement of the terms. The first line is the value of the price change at the new level of q_1. The second line is the negative of the value of the price change at the initial level of q_1. The last line is the value of the change in q_1 at the new price level. If a special condition referred to as "weak complementarity"—which is discussed next—is satisfied, this arrangement is useful and forms the basis for the travel cost method presented in Chap. 6.

2.2.2 Weak Complementarity

Suppose the compensated demand for Market Good 1 (x_1) depends on the level of q_1 in a marginally positive way; that is, the demand curve shifts out as q_1 increases. Further suppose that if consumption of this market good is zero, the marginal value for the change in q_1 is zero. Maler (1974) referred to this situation as weak complementarity. Now, turning back to the way that compensating surplus was rewritten in Eq. (2.17), suppose the change in price was from the original price level to the price that chokes off demand for this weakly complementary good. This choke price is designated as $\hat{p}_1$:

$$\begin{aligned} CS &= e\left(\hat{p}_1, P^0_{-1}, q^1_1, Q^0_{-1}, U^0\right) - e\left(p^0_1, P^0_{-1}, q^1_1, Q^0_{-1}, U^0\right) \\ &\quad - \left[e\left(\hat{p}_1, P^0_{-1}, q^0_1, Q^0_{-1}, U^0\right) - e\left(p^0_1, P^0_{-1}, q^0_1, Q^0_{-1}, U^0\right)\right] \\ &\quad + e\left(\hat{p}_1, P^0_{-1}, q^0_1, Q^0_{-1}, U^0\right) - e\left(\hat{p}^0_1, P^0_{-1}, q^1_1, Q^0_{-1}, U^0\right). \end{aligned} \tag{2.18}$$

By definition, demand for the weakly complementary good is zero at $\hat{p}_1$ and so the last line of Eq. (2.18) equals zero. Now the compensating surplus is simply the change in total consumer surplus for the weakly complementary good:

$$\begin{aligned} CS &= e\left(\hat{p}_1, P^0_{-1}, q^1_1, Q^0_{-1}, U^0\right) - e\left(p^0_1, P^0_{-1}, q^1_1, P^0_{-1}, U^0\right) \\ &\quad - \left[e\left(\hat{p}_1, P^0_{-1}, q^0_1, Q^0_{-1}, U^0\right) - e\left(p^0_1, P^0_{-1}, q^0_1, Q^0_{-1}, U^0\right)\right] \\ &\quad + \int_{\hat{p}^0_1}^{\hat{p}_1} x^h_1\left(s, P^0_{-1}, q^1_1, Q^0_{-1}, U^0\right) ds - \int_{p^0_1}^{\hat{p}_1} x^h_1\left(s, P^0_{-1}, q^0_1, Q^0_{-1}, U^0\right) ds. \end{aligned} \tag{2.19}$$

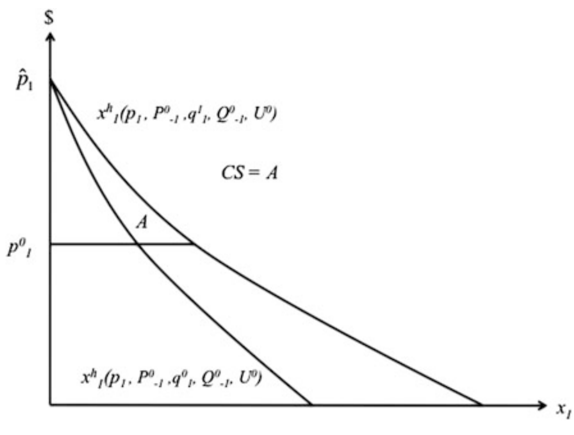


Fig. 2.3 Weak complementarity of market good x_1 and CS for a change in nonmarket good q_1

Weak complementarity is convenient because valuing the change in the non-market good is possible by valuing the change in consumer surplus from the weakly complementary good. Figure 2.3 graphically depicts compensating surplus for this weakly complementary good.

Consumption of several goods might need to be zero in order for the marginal value of q_1 to equal zero. An example is improving the water quality at two sites along a river. The value of improving water quality might be zero if trips to both sites were zero—a joint weak complementarity condition. These concepts are similar to those presented so far. The difference is the way the sequence of price changes is dealt with. The final line in the analog to (2.18) would still equal zero. However, there are multiple prices to consider. Consider a simple example of how the prices of two goods would need to be adjusted. Suppose that if demand for Market Good 1 and 2 is zero, then the marginal value for the change in q_1 equals zero. Compensating surplus is then given as follows. Similar to the earlier notation, $P^0_{-1,-2}$ is the price vector formed by removing the first and second elements of P^0:

$$\begin{aligned} CS = &\int_{\hat{p}_1^0}^{\hat{p}_1} x_1^h\left(s, P^0_{-1}, q_1^1, Q^0_{-1}, U^0\right) ds \\ &- \int_{\hat{p}_1^0}^{\hat{p}_1} x_1^h\left(s, P^0_{-1}, q_1^0, Q^0_{-1}, U^0\right) ds \\ &+ \int_{\hat{p}_2^0}^{\hat{p}_2} x_2^h\left(\hat{p}_1, s, P^0_{-1-2}, q_1^1, Q^0_{-1}, U^0\right) ds \\ &- \int_{p_2^0}^{\hat{p}_2} x_2^h\left(\hat{p}_1, s, P^0_{-1-2}, q_1^0, Q^0_{-1}, U^0\right) ds. \end{aligned} \tag{2.20}$$

Compensating surplus is given by the change in consumer surplus resulting from the increase in q_1 for the two goods. For the first good, the change in consumer surplus is conditioned on all of the other prices being held at the original level, P^0_{-1}. For the second good, the change in consumer surplus is conditioned on the choke price of the first good, $\hat{p}_1$, and the original price for the remaining market goods, P^0_{-1-2}. If there were a third good, the change in consumer surplus for the third good would be conditioned on the choke prices of Goods 1 and 2. This adjustment would be necessary for measuring changes in consumer surplus for any sequence of price changes—not just choke prices. The order of the price changes does not matter as long as the other prices are conditioned correctly (Braeutigam and Noll 1984).

Before moving on to inference for marginal values, two issues related to weak complementary goods should be mentioned. First, the analyst does not need to rule out a market price other than the choke price for which he or she obtains the condition that the marginal value of the market good is zero.[8] Any price that results in this condition allows the compensating surplus to be derived from the compensated market demands. The techniques discussed in this section will handle any such price. The second issue is the impact of incorrectly assuming weak complementarity. The last term that vanishes under weak complementarity will be positive if one incorrectly assumes weak complementarity for increases in the nonmarket good, and negative for decreases. The value inferred from the good that is incorrectly assumed to be weakly complementary will bound compensating surplus either below, for increases in the nonmarket good, or above for decreases.

2.2.3 Household Production Framework

A slightly different approach to that presented above is the household production framework. The household production framework is the basis for the defensive behavior approach to nonmarket valuation described in Chap. 8. Suppose the analyst is interested in determining the marginal value of a single nonmarket good q_j. The household production framework posits a production relationship between the consumption of goods x_p and q_j. The good produced in this process is a final product that the consumer values. Partition the vector X into $[X_{-p}, x_p]$, where x_p is a good produced by the individual according to the production process $x_p = f\left(I, q_j\right)$. I is a marketed production input, X_{-p} is the vector of market goods consumed, p_{-p} is a vector of prices for the market goods, and p_I is the price for the marketed production input. Assuming that q_j enters the choice problem only through production of x_p, the utility maximization problem is

$$\max_{X_{-p},I} U\left(X_{-p}, x_p\right) \; s.t. \; p_{-p} \cdot X_{-p} + p_I \cdot I \leq y, q_j = q_j^0, x_p = f(I, q_i). \tag{2.21}$$

The necessary conditions for this maximization problem imply two important equations that involve the marginal value of additional income, λ, and the marginal value (virtual price) of additional q_j given by $p^v_{q_j}$, the object of interest. With knowledge of the marginal value, one can approximate the value for a discrete change by integrating over the marginal value similar to Eqs. (2.10) and (2.11):

$$\frac{\partial U}{\partial x_p} \cdot \frac{\partial f}{\partial I} = \lambda \cdot p_1 \qquad \frac{\partial U}{\partial x_p} \cdot \frac{\partial f}{\partial q} = p^v_{q_j} \lambda. \tag{2.22}$$

From these two equations, one can solve for the marginal value of q_j:

---

[8] An example is the case of weak substitutability provided in Feenberg and Mills (1980).

$$p_q^v = p_I \cdot \frac{\left(\frac{\partial f}{\partial q}\right)}{\left(\frac{\partial f}{\partial I}\right)}. \tag{2.23}$$

Thus, the marginal value of q_j can be derived from the price of the marketed production input and the marginal rate of transformation between the input and q_j. The desirable property of this technique is that there is no need to model preferences. Of course, the analyst still has to model the production process. Moving away from a single input and a single produced good quickly complicates the model. Preferences need to be modeled because marginal utilities come into play. Therefore, the analyst needs to model the production process *and* consumer preferences, which creates an additional layer to the basic framework that has been presented.

2.2.4 The Hedonic Concept

Some goods that are consumed can be viewed as bundles of attributes. For example, houses have distinguishing attributes such as square footage, number of bedrooms, location, and environmental attributes. Public land is an example of a publicly owned environmental good that provides open space that is accessible to all. Being close to open space is, for some, a valuable attribute. Holding all other characteristics of houses constant, houses closer to open space have higher sale prices. Given this price gradient, purchasers of homes can buy location relative to open space up to the point where the marginal cost of moving closer equals the marginal benefit.

Hence, there is an implicit market in this attribute because the home price varies by distance to open space. This concept underlies the hedonic nonmarket valuation technique described in Chap. 7. Other examples of attributes in the housing market are air quality, busy streets, and power lines. Environmental risk is an attribute of jobs, which are objects of choice that implicitly offer people the chance to trade off pay and on-the-job risk of injury or exposure to toxins. The important feature of the hedonic model is that an implicit market exists for attributes of goods, such as distance to open space or job risk, which are not explicitly traded in markets.[9]

In the case of the home purchase, the idea is that the consumer purchases environmental quality through the house. Utility still depends on the consumption of market goods X and nonmarket goods Q, but now certain aspects of Q can be thought of as being chosen. It is important to recognize levels of rationing. For example, the consumer does not individually purchase open space; thus, the quantity of Q is fixed. He or she can, however, purchase a home closer to the open

[9]The classic citations in this area are Griliches (1971), Rosen (1974), and Palmquist (1984).

space that is available. For the case of air quality, the quality gradient is fixed so far as the individual is concerned.

A resident of a large city cannot unilaterally affect this gradient, but he or she can choose where to live along the gradient. The resident can choose a house based on where it falls along the air quality gradient. He or she can care a great deal about the gradient itself in ways other than the choice of housing. For example, the resident can live near the beach, which has relatively good air quality, and yet be subjected to really poor air quality at work downtown. Similarly, the resident can locate near the North Boulder open space and yet care a great deal about whether Boulder County purchases another tract of land in South Boulder. The point here is that Q can enter one's utility for life at home and also enter separately in the utility function for other purposes.

The basic approach to the hedonic method is that the house is really a bundle of attributes. Because other people also care about these attributes, they are scarce and valuable. Although the consumer pays a bundled price for the house, the price buys the package of individual attributes. A way to model things on the consumer side is to partition both market goods, $X = [X_1, X_2]$, and nonmarket goods, $Q = [Q_1, Q_2]$. The second vector in both the market and nonmarket goods partitions are those attributes selected through the housing purchase. The total price of the house is a function of these attributes, $p_h(X_2, Q_2)$. The maximization problem follows:

$$\begin{aligned} &\max_{X,Q_2} U(X_1, X_2, Q_1, Q_2) \\ s.t \quad &p_1 \cdot X_1 + p_h(X_2, Q_2) \leq y, \quad Q_1 = Q_1^0. \end{aligned} \tag{2.24}$$

The important feature is that the consumer chooses the levels of Q_2 through the house purchase up to the point where the marginal benefit equals marginal cost. In particular, the marginal rates of substitution for elements in Q_2 and X_2 are equal to the relative marginal costs, i.e., prices:

$$\begin{aligned} &\frac{\left(\frac{\partial U}{\partial q_j}\right)}{\left(\frac{\partial U}{\partial x_j}\right)} = \frac{\left(\frac{\partial p_h}{\partial q_j}\right)}{\left(\frac{\partial p_h}{\partial x_j}\right)} \quad q_j \in Q_2, \quad x_j \in X_2 \\ &\frac{\left(\frac{\partial U}{\partial q_j}\right)}{\left(\frac{\partial U}{\partial x_j}\right)} = \frac{\left(\frac{\partial p_h}{\partial q_j}\right)}{p_i} \quad q_j \in Q_2, \quad x_j \in X_2. \end{aligned} \tag{2.25}$$

As in the case for market goods, the combined marginal substitution relationships conceptually yield a marginal substitution curve, referred to as the bid function for the individual. Conversely, sellers are typically trying to get the most money possible for their houses. The price function, $p_h(X_2, Q_2)$, is the resulting equilibrium from the interaction of buyers and sellers. Estimating the price function using demand provides information on the marginal values of Q_2. Additional structure that is discussed in Chap. 7 facilitates estimation of the demand functions that can then be used to value nonmarginal changes.

2.2.5 When Markets Will Not Do

The concepts outlined in the earlier sections involve the use of observable market behavior to infer either the marginal value of nonmarket goods or the value for a discrete change in the nonmarket goods. All of these methods require an identifiable link between the nonmarket goods and some subset of the market goods. Furthermore, there also must be sufficient variation in the prices of the market goods and the quantities or qualities of the nonmarket goods accompanying the observed transactions to be able to statistically identify these relationships. The concepts outlined in the earlier sections form the basis for revealed preference nonmarket valuation techniques described in Chaps. 6, “Travel Cost”; 7, “Hedonics”; and 8, “Defensive Behavior and Damage Cost Methods.”

Using market data to infer the value of a nonmarket good requires that values can only be inferred for individuals who used the nonmarket good, but there are cases when the demand link is unidentifiable for some individuals. A lack of identifiable link for some people does not mean they do not value the nonmarket good. Value for these individuals for whom there is no identifiable or estimable link is referred to as nonuse value or passive-use value. Passive-use value is the legal term used by the U.S. Federal Court of Appeals in an influential court case, Ohio v. U.S. Department of the Interior, which gave legal standing to the concept. Drawing on earlier work from Carson et al. (1999), a brief overview of how this concept evolved follows.

In a highly influential article, Krutilla (1967) suggested that revealed preference techniques might not accurately measure societal values. The strength of his argument came through examples; the paper provides no theory. Using unique resources such as the Grand Canyon, and considering irreversible changes, Krutilla (1967) made a number of important points.[10]

First, demand for the environment has dynamic characteristics that imply value for potential use, though not current use, and trends for future users need to be explicitly recognized in order to adequately preserve natural areas.[11]

Second, some individuals may value the environment for its mere existence. Krutilla (1967, footnote 7, p. 779) gave the example of the “spiritual descendants of John Muir, the current members of the Sierra Club, the Wilderness Society, National Wildlife Federation, Audubon Society and others to whom the loss of a

[10]Cicchetti and Wilde (1992) had contended that Krutilla’s (1967) arguments, and hence passive-use value, only apply to highly unique resources. However, Krutilla (Footnote 5, p. 778) noted that “Uniqueness need not be absolute for the following arguments to hold.”

[11]In discussing trends, Krutilla (1967) gave the example of the evolution from a family that car camps to a new generation of backpackers, canoe cruisers, and cross-country skiers.

species or a disfigurement of a scenic area causes acute distress and a sense of genuine relative impoverishment."

Third, the bequest of natural areas to future generations may be a motive for current nonusers to value preservation, particularly because given the dynamic characteristics mentioned previously, preserving natural areas effectively provides an estate of appreciating assets.

These examples obviously struck a chord with many economists. Methods and techniques were developed to formally describe the phenomena mentioned by Krutilla (1967) and to measure the associated economic value.[12]

Measuring passive-use values and using them in policy analysis—particularly natural resource damage assessments—has been controversial. Much of the problem stems from the fact that passive-use values, by implied definition, cannot be measured from market demand data. Economics, as a discipline, places considerable emphasis on drawing inferences regarding preferences from revealed actions in markets. However, stated preference methods such as contingent valuation (Chap. 4) and choice experiments (Chap. 5) are the only viable alternatives for measuring passive-use values. These stated preference methods draw inference from carefully designed scenarios of trade-offs that people are asked to evaluate in survey settings. From the trade-off responses, we learn about the preferences of individuals who hold passive-use values.

Some economists are skeptical about passive-use values in economic analysis, and the skepticism occurs on two levels. The first level involves the idea of whether or not passive-use values even exist. The second level involves the measurement of passive-use values because of the need to use stated preference techniques. To completely dismiss passive-use values is an extreme position and does not hold up to scrutiny because nonusers of areas like the Arctic Wildlife Refuge or the Amazon rain forest frequently lobby decision-makers to preserve these areas and spend money and other resources in the process. The latter concern is based on empirical observations that have been published in the literature.[13]

The remainder of this section will discuss how passive-use values have been viewed conceptually. While the discussion will focus on compensating surplus, the issues also apply to equivalent surplus. Recall the decomposition of compensating surplus into the value of a price change to the choke price and the value of the quantity change at the higher price level. Weak complementarity called for the final term to equal zero in Eq. (2.18). McConnell (1983) and Freeman (1993) defined passive-use value as this last term:

[12]The terms "option value," "preservation value," "stewardship value," "bequest value," "inherent value," "intrinsic value," "vicarious consumption value," and "intangible value" have been used to describe passive-use values. Carson et al. (1999) noted that these are motivations rather than distinct values.

[13]See Carson (2012), Portney (1994), Hanemann (1994), and Diamond and Hausman (1994).

$$\begin{aligned} CS = {} & e\left(\hat{p}_1, p^0_{-1}, q^1_1, Q^0_{-1}, U^0\right) - e\left(p^0_1, p^0_{-1}, q^1_1, Q^0_{-1}, U^0\right) \\ & - \left[e\left(\hat{p}_1, p^0_{-1}, q^0_1, Q^0_{-1}, U^0\right) - e\left(p^0_1, p^0_{-1}, q^0_1, Q^0_{-1}, U^0\right)\right] \\ & + PUV. \end{aligned} \tag{2.26}$$

This definition does not have much practical appeal because we could choose any good that is not a necessary good, measure the value from the first two lines of (2.26), and end up with a measure of passive-use value. Because one could do this for each good that is not a necessary good or any combination of goods in this category, multiple measures of passive-use value could be derived.

Another conceptual definition was suggested by Hanemann (1995) with a specific form of utility in mind, $U(X, Q) = T[g(X, Q), Q]$. This functional form suggests that choices of market goods will be influenced by Q, and so market demand data could reveal the part of the relationship involving $g(X, Q)$ but not the part where Q enters directly.[14] Hanemann (1995) defined passive-use value and use value according to the following two identities:

$$T\left[g\left(X\left(P, Q^0, y - PUV\right), Q^0\right), Q^1\right] = T\left[g\left(X\left(P, Q^0, y\right), Q^0\right), Q^0\right]; \text{and} \tag{2.27}$$

$$T\left[g\left(X\left(P, Q^1, y - PUV - UV\right), Q^1\right), Q^1\right] = T\left[g\left(X\left(P, Q^0, y\right), Q^0\right), Q^0\right]. \tag{2.28}$$

The definitions implied by (2.26) and by (2.27) together with (2.28) decompose compensating surplus into two parts for which the sum of the parts equals the whole. Intuitively, Hanemann's (1995) definition works in reverse of the decomposition in (2.26). Because the same preferences can be defined differently, a passive-use value is a somewhat tenuous theoretical concept.[15] Furthermore, neither definition is easy to implement because the first decomposition requires one to choose the market goods for which demand is choked. Using separate measurement, it is difficult if not impossible to elicit either of these concepts from subjects in a stated preference study.

Carson et al. (1999) provided a definition based on methodological considerations. "Passive-use values are those portions of total value that are unobtainable using indirect measurement techniques which rely on observed market behavior" (p. 100).[16]This definition was conceived with the researcher in mind as opposed to a theoretical foundation. Revealed preference techniques can miss portions of value because of the form of preferences such as those used in the Hanemann (1995) definition. Analysts typically want to measure compensating or equivalent surplus,

[14]The special case where $g(X, Q) = g(X)$ has been referred to as "the hopeless case" because the ordinary demands are independent of the levels of Q, leaving no hope for recovering the value of Q from demand data.

[15]Dividing passive-use value into bequest value, existence value, and the like will provide similarly inconclusive results. The decompositions will not be unique.

[16]Maler et al. (1994) similarly defined use values as those values that rely on observed market behavior for inference.

also referred to as total value in this literature, but are less concerned with separate estimates of individual use and passive-use elements of the total value. The important social issue is the need to incorporate the values of all those who value the nonmarket good. To do so requires analysts, at times, to turn to stated preference techniques if they believe that passive-use values are likely to be decisive. Similarly, sometimes the use-value component alone can sufficiently inform decisions and allow analysts to rely on revealed behavior that some view as more credible. It is important to recognize that separating out use and passive-use values at the individual level is quite difficult and sometimes impractical because preferences along these two dimensions frequently interact.

2.3 Nonmarket Values in a Dynamic Environment

Most models of intertemporal choice in economics assume that utility across time periods is represented by a sum of utility functions from each of the time periods. This sum involves a time preference component that is typically assumed to be the discount factor, $\gamma = 1/(1+r)$:

$$U = \sum_{t}^{T} \gamma^t u(X_t, Q_t). \tag{2.29}$$

Utility in each period depends on market goods X_t and nonmarket goods Q_t. The time horizon, T, can be either finite or infinite. The analog to the earlier problem is that the consumer still allocates income toward the purchase of market goods, but now total income is in present value form, $Y = \sum \gamma^t y_t$, where y_t is income in period t. A simple time separable model such as this can be used to extend the earlier concepts of value developed for a single period into a dynamic framework. Assume that X_t is a composite good consisting of expenditures on the market goods in period t. Thus, expenditures on market goods and levels of nonmarket goods (Q_t) exist in each period. The important feature of this model is that the individual efficiently allocates income between periods. That is to say, the marginal benefit of spending on market goods in each period is equated in present value terms:

$$\frac{\partial u(X_0, Q_0)}{\partial X} = \gamma^t \frac{\partial u(X_t, Q_t)}{\partial X}. \tag{2.30}$$

This condition must hold for all t under optimal income allocation. The consideration is what a marginal change in Q_t is worth in the current period. The marginal value for the change will be given by $p_t^v = (\partial u(X_t, Q_t)/\partial Q)/(\partial u(X_t, Q_t)/\partial X)$. By (2.30), the value today for the marginal change in the future will simply be given by $\gamma^t p_t^v$. Thus, the margin value of Q in the dynamic model is simply the discounted value of the marginal value in the respective period.

For discrete changes, the analyst would like the total amount of income today that the consumer is willing to give up for some change in the sequence of nonmarket goods, $\{Q_t\}$. Brackets are used because the levels of nonmarket goods for T periods must be monitored, and T may be infinite. Assuming one is interested in a compensating measure of welfare, the logical extension from the marginal case is to use the present value discounted stream of compensating surplus in each period as the welfare measure. This generalization meets the needs provided that the allocation of income is unaffected by the sequence of nonmarket goods, $\{Q_t\}$. However, when income allocation is affected by this sequence, the proposed welfare measure, the present value discounted compensating surplus in each period, essentially values the changes in the sequence while imposing that the income allocation across periods is fixed. Thus, for increases in nonmarket goods, the present value of the compensating surpluses will underestimate the desired welfare measure, and the present value of equivalent surpluses will overstate the amount. The reasoning is that for both cases, the ability to reallocate income is worth money. For the compensating measure, one would pay for this flexibility over the restricted case measured by the present value of the compensating surpluses from each period. For equivalent surplus, the ability to reallocate income makes giving up the change in $\{Q_t\}$ not as bad. For decreases in $\{Q_t\}$, the opposite is true in both cases.

Practically speaking, the standard practice is to estimate the periodic benefits and then discount them. The choice of the discount rate is a sensitive issue that will not be addressed here.[17] Because the estimation is of future benefits in today's dollars, the appropriate discount rate should not include an inflationary component.

2.3.1 Values in an Uncertain World

A great amount of uncertainty exists regarding our willingness to trade money for nonmarket goods. For example, the levels of nonmarket goods provided by a policy may be uncertain, prices of market goods that will occur once the policy is implemented may be uncertain, and the direct cost if the policy is enacted may be uncertain. Policies can affect the distributions of all these random variables. The question then becomes one of how to extend the welfare measures developed in the previous section to cases of uncertainty.

Exclusively consider uncertainty regarding the distribution of Q, assuming away time.[18] Q can accommodate things as different as the total amount of open space that will be purchased by a bond initiative or the level of environmental risk associated with living or working in a given area. In relation to the earlier models,

---

[17]For examples, see Fisher and Krutilla (1975), Horowitz (1996), Porter (1982), and Schelling (1997).

[18]Time is an important dimension, and uncertainty transcends time. However, there is not enough space to cover time and uncertainty together.

one now assumes that individuals allocate income toward the purchase of market goods according to expected utility maximization:

$$\max \mathrm{E}_Q[U(X,Q)] \quad s.t. \quad P \cdot X \le y. \tag{2.31}$$

Here, the allocation of income depends on the distribution of Q, which involves different possible levels instead of a particular level. The distribution of Q can be discrete or continuous. The maximized expected utility depends on the prices of the market goods, income, and the probability distribution of Q. Values that influence policy choices are now dependent on the distribution associated with Q. Letting F denote the probability distribution of Q, maximized expected utility is then given by an indirect utility function, $v^E(P,y,F)$.[19] The central concept is option price. Option price is defined as the amount of money, a reduction in income in this example, that makes the individual indifferent between the status quo level of expected utility and the new expected utility under the changed distribution:

$$v^E\left(P, y - OP, F^1\right) = v^E\left(P, y, F^0\right). \tag{2.32}$$

Here, OP is the measure of compensating surplus under uncertainty. In cases such as bond issues for the provision of open space, residents typically pay some single, specified amount over time. The amount of open space that will actually be purchased is uncertain. In this case, option price is a very close analog to compensating surplus from the open space example in Sect. 2.1.5. In fact, contingent valuation surveys generally measure option price because some uncertainty almost always exists. Other important concepts involving environmental valuation and uncertainty are not covered here.[20]

2.3.2 Averting Expenditures

This section develops the broad conceptual framework for using averting expenditures as a means to value nonmarket goods—a topic that is taken up in detail in Chap. 8. When facing environmental risks, individuals may independently undertake costly risk reductions. Examples include the purchase of bottled water and purchasing air bags for the car, to name a few. Because individuals spend money to provide a more favorable probability distribution of the nonmarket good, averting expenditures offers an avenue for inferring the value of collective policies that affect the distribution. The idea here is that the probability distribution can be favorably

[19]In accordance with standard probability theory, F consists of a sample space of outcomes and a probability law for all subsets of the sample space that satisfies the properties of a probability measure.

[20]Influential papers in this area include Graham (1981), Weisbrod (1964), Schmalensee (1972), and Arrow and Fisher (1974). Freeman (1993) provided a fairly comprehensive overview.

affected through individual inputs as well as collective inputs. Let E_I denote the individual's expenditure dedicated toward individual improvement of the distribution of Q, and let E_G denote the government's expenditure dedicated toward improving this distribution. Now the individual chooses both the level of market expenditures and the level of E_I subject to E_G. As in the previous section, F is the probability distribution of Q. At the optimal level of E_I, the indirect expected utility function becomes $v^E(P, y, F(E_I, E_G))$. A necessary condition for optimization is that the marginal benefit of more E_I equals the marginal utility of additional income:

$$\frac{\partial v^E}{\partial F}\frac{\partial F}{\partial E_I} = \lambda. \tag{2.33}$$

The marginal value of additional government expenditure dedicated toward improving the distribution of Q, denoted p_G^v, can be represented as the marginal utility of the expenditure function divided by the marginal utility of income:

$$p_G^v = \frac{\partial v^E}{\partial F}\frac{\partial F}{\partial E_I}\frac{1}{\lambda}. \tag{2.34}$$

From (2.33) and (2.34), one can solve for the marginal value of E_G. The way in which E_I enters the problem, the marginal value of E_G reduces to what is similar to the marginal rate of transformation for inputs:

$$p_G^v = \frac{\frac{\partial v^E}{\partial F}\frac{\partial F}{\partial E_G}}{\frac{\partial v^E}{\partial F}\frac{\partial F}{\partial E_I}} = \frac{\frac{\partial F}{\partial E_G}}{\frac{\partial F}{\partial E_I}} \tag{2.35}$$

In this case, one only needs to understand the relative production possibilities between private and public expenditures. This technique is conceptually similar to the household production framework. As with the household production framework, if expenditures made toward improving the probability distribution also affect other aspects of utility, the marginal value expression is more complicated than (2.35).

2.3.3 Welfare in Discrete-Choice Random Utility Models

Chapters 4, 5, 6, and 8 present discrete-choice random utility models that can be applied either in stated preference or revealed preference settings. Discrete-choice random utility models seek to describe choices over distinct alternatives that frequently vary by common attributes. For example, sport fishermen choose among competing fishing sites that are distinguished by distance from home as well as catch rates. Or choice experimental subjects may choose among alternative policies that are distinguished by policy attributes such as acres of wilderness areas preserved, the number of species affected, and differences in household taxes.

The basic idea behind these models is that people receive utility from a given alternative that depends on a set of observable characteristics and other characteristics that are not observed by the researcher. The indirect utility of alternative j, conditional on having made that choice, is a function of a vector of observable characteristics Q_j, an unobservable random component ε_j, income y, and cost of that alternative $p_j : v_j = v(p_j, Q_j, \varepsilon_j, y)$. Specification of the functional form of the conditional indirect utility function and assumptions made regarding the probability distribution of the unobservable random component facilitate modeling the probability of choosing available alternatives that, in turn, provides estimates of conditional indirect utility function parameters. With regard to welfare analysis, the concepts are identical in spirit to those discussed above, though some consideration must be given to the unobservable random component associated with each alternative.

In the recreational demand setting, one could consider a change in the characteristics of, say, Site 1. The analysis starts with an initial set of vectors of observable characteristics for all J sites, $\{Q_1^0, Q_2^0, \ldots, Q_J^0\}$. The policy being considered changes the characteristics of one site, Site 1 here, $\{Q_1^1, Q_2^0, \ldots, Q_J^0\}$. With regard to welfare measures, of interest are the amount of income adjustment that would make an individual indifference between the initial set of characteristics and the new set of characteristics provided by the policy with the appropriate income adjustments, e.g., subtracting C along with the policy or adding E while forgoing the policy.

This framework allows the estimation of marginal values but is not designed to reveal specific site or alternative choices with changed conditions. That is, the random component in conditional indirect utility does not allow us to say with certainty which site will be chosen on the next trip with either the initial set of vectors or the new vector. Consider the optimal choice that the angler makes presented in (2.36). With regard to notation, $\{p\}$, $\{Q\}$, $\{\varepsilon\}$ are, respectively, the set of prices for each site, the set of observable attributes for each site, and the set of unobservable utility components for each site that is unknown to the researcher:

$$v^*(\{p\}, \{Q\}, \{\varepsilon\}, y) = \max_{j \in J} v\big(p_j, Q_j, \varepsilon_j, y\big). \tag{2.36}$$

In Sect. 2.3.1 on values in an uncertain world, individuals face uncertainty over Q. In this model, it is the researcher who faces uncertainty over the random utility components. In order to derive welfare measures, he or she considers the expected maximum utility as the value function, and the compensating and equivalent measures are as follows. Here, expectation is with regard to the unobservable error components, and the income adjustments are the same regardless of the site that is actually chosen:

$$E_{\varepsilon}\left[v^*\left(\{p\},\{Q^0\},\{\varepsilon\},y\right)\right] = E_{\varepsilon}\left[v^*\left(\{p\},\{Q^1\},\{\varepsilon\},y-C\right)\right]; \tag{2.37}$$

$$E_{\varepsilon}\left[v^*\left(\{p\},\{Q^0\},\{\varepsilon\},y+E\right)\right] = E_{\varepsilon}\left[v^*\left(\{p\},\{Q^1\},\{\varepsilon\},y\right)\right]. \tag{2.38}$$

The difficulty in analytically deriving these measures depends on the form of conditional indirect utility, the assumed distribution of the unobservable error components, and the number of choices available to agents.[21] In some cases—such as binary choice questions in contingent valuation—solutions to these equations are straightforward, while in other cases, solutions are analytically intractable. It is important to recognize that in all cases, the basic principles are consistent with measures presented in Sect. 2.2.

2.4 Parting Thoughts

All of the models presented in this chapter are based on the assumptions that individuals understand their preferences and make choices so as to maximize their welfare. Even under optimal conditions, inferring values for nonmarket goods is difficult and requires thoughtful analysis. The nonmarket valuation practitioner needs to understand these concepts before heading into the field; to do otherwise could prove costly. Estimating credible welfare measures requires careful attention to the basic theoretical concepts so that value estimates meaningfully and clearly support decision-making. There has never been a shortage of critics of welfare economics—from inside or outside the profession. Nonmarket valuation researchers are on the cutting edge of these conceptual issues, a necessary trend that will undoubtedly continue.

References

Arrow, K. J. & Fisher, A. C. (1974). Environmental preservation, uncertainty, and irreversibility. Quarterly Journal of Economics, 88, 312-319.

Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. Journal of Political Economy, 101, 385-409.

Bockstael, N. E. & McConnell, K. E. (2007). Environmental and resource valuation with revealed preferences: A theoretical guide to empirical models. Vol. 7 of I. J. Bateman (Ed.), The Economics of Non-Market Goods and Resources series. Dordrecht, The Netherlands: Springer.

Braeutigam, R. R. & Noll, R. G. (1984). The regulation of surface freight transportation: The welfare effects revisited. Review of Economics and Statistics, 66, 80-87.

Carson, R. T. (2012). Contingent valuation: A practical alternative when prices aren't available. Journal of Economic Perspectives, 26 (4), 27-42.

[21]In Chap. 5 of their book, Bockstael and McConnell (2007) provided an excellent overview of welfare measures in discrete-choice models.

Carson, R. T., Flores, N. E. & Mitchell, R. C. (1999). The theory and measurement of passive-use value. In I. J. Bateman & K. G. Willis (Eds.), Valuing environmental preferences: Theory and practice of the contingent valuation method in the US, EU, and developing countries (pp. 97-130). Oxford, United Kingdom: Oxford University Press.

Cicchetti, C. J. & Wilde, L. L. (1992). Uniqueness, irreversibility, and the theory of nonuse values. American Journal of Agricultural Economics, 74, 1121-1125.

Diamond, P. A. & Hausman, J. A. (1994). Contingent valuation: Is some number better than no number? Journal of Economic Perspectives, 8 (4), 45-64.

Feenberg, D. &. Mills, E. S. (1980). Measuring the benefits of water pollution abatement. New York: Academic Press.

Fisher, A. C. & Krutilla, J. V. (1975). Resource conservation, environmental preservation, and the rate of discount. Quarterly Journal of Economics, 89, 358-370.

Freeman, M. A. III. (1993). The measurement of environmental and resource values: Theory and methods. Washington, DC: RFF Press.

Graham, D. A. (1981). Cost-benefit analysis under uncertainty. The American Economic Review, 71, 715-725.

Griliches, Z. (1971). Price indexes and quality change: Studies in new methods of measurement. Cambridge, MA: Harvard University Press.

Hanemann, W. M. (1991). Willingness to pay and willingness to accept: How much can they differ? The American Economic Review, 81, 635-647.

Hanemann, W. M. (1994). Valuing the environment through contingent valuation. Journal of Economic Perspectives, 8 (4), 19-43.

Hanemann, W. M. (1995). Contingent valuation and economics. In K. G. Willis & J. T. Corkindale (Eds.), Environmental valuation: New perspectives (pp. 79-117). Wallingford, Oxfordshire, United Kingdom: CAB International.

Hanemann, W. M. (1999). Neo-classical economic theory and contingent valuation. In I. J. Bateman & K. G. Willis (Eds.), Valuing environmental preferences: Theory and practice of the contingent valuation method in the US, EU, and developing countries (pp. 42-96). Oxford, United Kingdom: Oxford University Press.

Hanley, N., Shogren, J. F. & White, B. (Eds.). (1997). Environmental economics: In theory and practice. New York: Oxford University Press.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. The American Economic Review, 71, 662-676.

Hicks, J. R. (1943). The four consumer's surpluses. The Review of Economic Studies, 31-41.

Horowitz, J. K. (1996). Environmental policy under a non-market discount rate. Ecological Economics, 16, 73-78.

Kreps, D. M. (1990). A course in microeconomic theory. Princeton, NJ: Princeton University Press.

Krutilla, J. V. (1967). Conservation reconsidered. The American Economic Review, 57, 777-786.

Maler, K.-G. (1974). Environmental economics: A theoretical inquiry. Baltimore, MD: Johns Hopkins Press for Resources for the Future.

Maler, K.-G., Gren, I.-M. & Folke, C. (1994). Multiple use of environmental resources: A household production function approach to valuing natural capital. In A. Jansson, M. Hammer, C. Folke & R. Costanza (Eds.), Investing in natural capital: The ecological economics approach to sustainability (pp. 233-249). Washington, DC: Island Press.

McConnell, K. E. (1983). Existence and bequest value. In R. D. Rowe & L. G. Chestnut (Eds.), Managing air quality and scenic resources at national parks and wilderness areas (pp. 254-264). Boulder, CO: Westview Press.

Palmquist, R. B. (1984). Estimating the demand for the characteristics of housing. Review of Economics and Statistics, 66, 394-404.

Porter, R. C. (1982). The new approach to wilderness preservation through benefit-cost analysis. Journal of Environmental Economics and Management, 9, 59-80.

Portney, P. R. (1994). The contingent valuation debate: Why economists should care. Journal of Economic Perspectives, 8 (4), 3-17.

Randall, A. & Stoll, J. R. (1980). Consumer's surplus in commodity space. The American Economic Review, 70, 449-455.
Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy, 82, 34-55.
Schelling, T. C. (1997). Intergenerational discounting. In T. H. Tietenberg (Ed.), The economics of global warming (pp. 607-613). Cheltenham, United Kingdom: Edward Elgar.
Schmalensee, R. (1972). Option demand and consumer's surplus: Valuing price changes under uncertainty. The American Economic Review, 62, 813-824.
Varian, H. R. (1992). Microeconomic analysis (3rd ed.). New York: W. W. Norton.
Vartia, Y. O. (1983). Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions. Econometrica, 51, 79-98.
Weisbrod, B. A. (1964). Collective-consumption services of individual-consumption goods. Quarterly Journal of Economics, 78, 471-477.
Willig, R. D. (1976). Consumer's surplus without apology. The American Economic Review, 66, 589-597.

Chapter 3
Collecting Nonmarket Valuation Data

Patricia A. Champ

Abstract This chapter describes how to collect primary and secondary data for nonmarket valuation studies. The bulk of this chapter offers guidance on how to design and implement a high-quality nonmarket valuation survey. Understanding the full data collection process will also help in evaluating the quality of data that someone else collected (i.e., secondary data). As there are not standard operating procedures for collecting nonmarket valuation, this chapter highlights the issues that should be considered in each step of the data collection process from sampling to questionnaire design to administering the questionnaire. While high-quality data that reflect individual preferences will not ensure the reliability or validity of estimated nonmarket values, quality data are a prerequisite for reliable and valid nonmarket measures.

Keywords Primary data · Secondary data · Nonmarket valuation survey · Sampling procedures · Study population · Sample frame · Survey mode · Survey documentation

The first two chapters in this book describe how individual choices are the foundation of nonmarket values. While that concept is relatively straightforward, one might wonder how we know about the choices an individual has made or would make. In other words, where do the data come from that reflect individuals' choices for nonmarket goods?

Sometimes market data can be used to infer choices about nonmarket goods. For example, I paid a premium price for the lot where I built my house. I did this because I liked that the lot backed up to open space. I could have purchased an almost identical lot in the same neighborhood that did not border the open space at a lower price. One could make an inference about my preference for open space from that market transaction. If there are many such transactions, one could use the data on lot sale prices in my neighborhood to infer the value of the open space behind my house.

P.A. Champ (✉)
U.S. Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA
e-mail: pchamp@fs.fed.us

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_3

More often, choices over environmental goods are not observable in market transactions. For example, an individual could care deeply and be willing to pay for the preservation of an endangered tiger species. However, there might not be a tiger preservation program that provides an opportunity to do that. In that situation, the researcher must go directly to individuals and ask what choices they would make about a particular nonmarket good or service if they had a chance to make such choices. Direct questioning about choices is done in surveys.

Surveys are prevalent. The current survey-crazed environment makes it difficult for potential respondents to know whether a survey is legitimate or a marketing ploy. Therefore, good survey design and scientific practice are more important than ever.

The bulk of this chapter offers guidance on how to design and implement a high-quality nonmarket valuation survey. Understanding the full data collection process will also help in evaluating the quality of data that someone else collected (i.e., secondary data). While high-quality data that reflect individual preferences will not ensure the reliability or validity of estimated nonmarket values, and quality data are a prerequisite for reliable and valid nonmarket measures. As the saying goes, "garbage in, garbage out." If data are not of sufficient quality, the corresponding nonmarket valuation estimates will be suspect.

While standard operating procedures do not exist for the collection of survey data, this chapter focuses on considerations and tradeoffs that are made in each step of the design and in administration of a nonmarket valuation survey. Throughout the chapter, the term "survey" refers to both the methods and the materials used to collect data. "Questionnaire" refers to the set of questions on the survey form. Following the discussion on developing and implementing a survey, the use of secondary data is discussed.

3.1 The Objective of the Nonmarket Valuation Study

Before worrying about data, the researcher must clarify the objectives of the study. The overall objective might be, for example, to inform policy, to assess the magnitude of damage to a natural resource, or to improve a nonmarket valuation method. Given the broad study objectives, the researcher must specify the economic question to be answered and decide whether equivalent surplus or compensating surplus will be measured. The economic question should be stated using the framework described in Chap. 2. If the study addresses a research question, the researcher should also develop testable hypotheses. The study objectives and economic questions will determine the most appropriate nonmarket valuation technique. In turn, the technique will determine what data are needed to estimate values.

In Chap. 1, Table 1.1 summarizes the broad steps in the valuation process. Data collection for nonmarket valuation falls under Step 5. The first order of business is to ascertain whether existing data can be used or whether it is necessary to collect new data. If new (i.e., primary) data need to be collected using a survey, decisions about the sampling techniques, mode of administration, and questionnaire content will be made in light of the overall study objectives and the specific economic questions to be addressed. If secondary data can be used to estimate nonmarket values, the researcher must decide where to get the data. As secondary data were once primary data, this chapter (in Sect. 3.2) largely focuses on conducting surveys to collect primary nonmarket valuation data. Section 3.3 then discusses secondary data sources.

3.2 Nonmarket Valuation Survey Methodology

The rapid changes in communication technology have had a huge impact on survey research. In the past, surveys were typically implemented start-to-finish using just one approach such as the telephone, the mail, or an in-person interviewer. Those days are gone.

It is now common for survey data to be collected using multiple approaches. For example, potential respondents can initially be contacted with a letter in the mail and given a choice of completing a paper or online questionnaire. A telephone call can then be used to remind nonrespondents to complete the questionnaire. Surveys implemented with more than one mode are referred to as mixed-mode surveys. As technology continues to change, new possibilities for collecting survey data will emerge. This chapter focuses on fundamental considerations that are likely to endure in the future for implementing a survey.

Before covering development and implementation of a nonmarket valuation survey, this section describes the process for deciding who will be invited to complete the survey. The survey objective will influence this decision. While some nonmarket surveys are conducted to provide values that can be generalized to the population of interest, there are other situations where the survey results are not intended to represent a broader population. For example, the economic experiments described in Chap. 10 may involve administering surveys to convenience samples drawn from student populations. The survey could be developed to allow for comparisons of values across similar groups as a test of economic theory or for tests of methodological issues related to nonmarket valuation. In both of these situations, the steps in development of the questionnaire and other survey materials will be similar. However, the situations will differ in the recruitment of survey respondents or in sampling procedures. In practice, the researcher makes decisions about sampling procedures and design of survey materials at the same time, as sampling options inform design and vice versa. However for the sake of exposition, survey sampling is described before survey design.

3.2.1 Sampling Procedures

There are two general approaches to sampling: probability sampling and nonprobability sampling (Fig. 3.1). Probability sampling involves a random selection process for inclusion of units in the sample where each unit of the study population has a known nonzero chance of being selected. In contrast, with nonprobability sampling, each individual in the population does not have a known nonzero probability of being

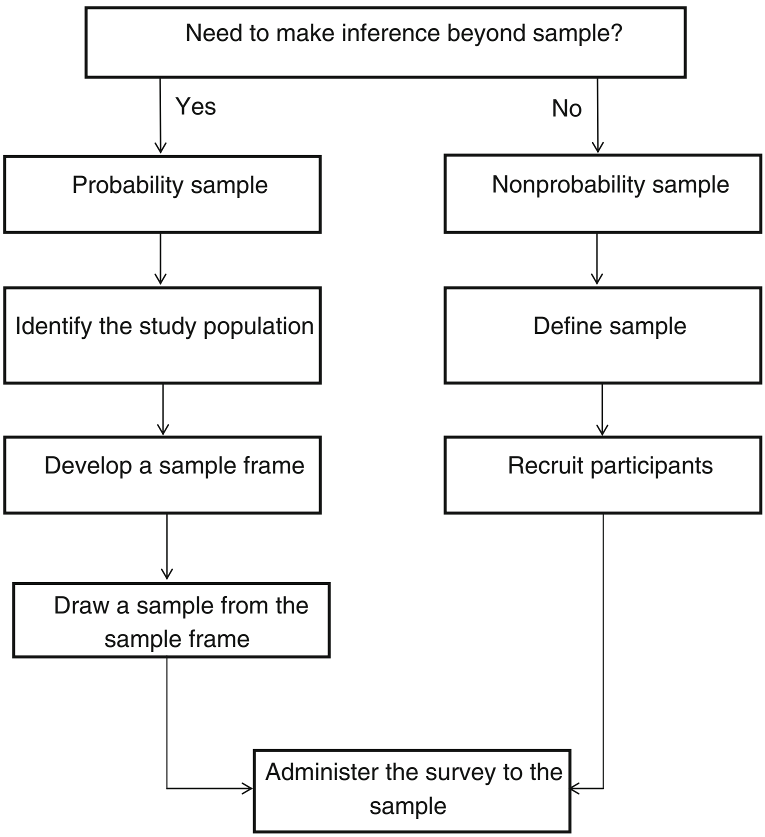


Fig. 3.1 Sampling

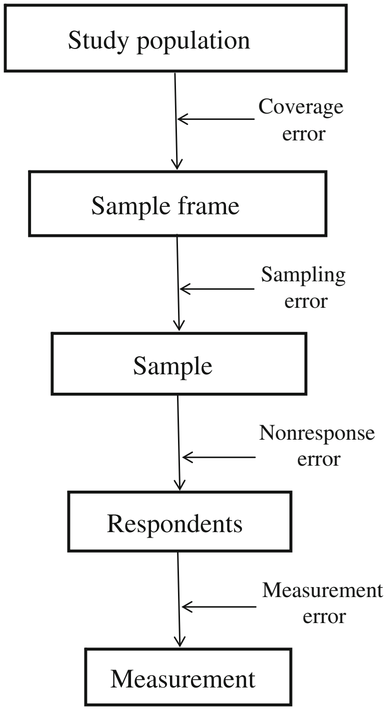
chosen for the sample. Examples of nonprobability sampling include using students in a class and recruiting subjects in a shopping mall. Nonprobability sampling is best-suited for studies that do not require generalization from the survey sample to a broader population. Such samples are frequently used to investigate methodological questions such as how responses are affected by survey design features.

The next three sections describe the steps associated with developing a probability-based sample. The sample design specifies the population of interest (study population), the sample frame, and the techniques for drawing a sample from the sample frame. Figure 3.1 outlines the steps in the sample design, while Fig. 3.2 identifies the potential sources of error that can arise during the survey process. Section 3.2.1.4 covers approaches for nonprobability sampling.

3.2.1.1 Study Population

Defining the study population begins with determining whose values are relevant. Consider a city interested in providing free Wi-Fi to a particular neighborhood. If the city would like to measure the benefits to neighborhood residents of providing

Fig. 3.2 Potential sources of error



free Wi-Fi, the appropriate study population is all residents who live within the neighborhood. This study population includes residents of the neighborhood who currently use Wi-Fi as well as those who do not. Including all neighborhood residents allows for measurement of the benefits to both current and potential Wi-Fi users.

However, the benefits of free Wi-Fi also can accrue to individuals who do not live in the neighborhood but visit the neighborhood for work or recreation. Further, it is possible that some individuals who neither live in nor visit the neighborhood could benefit from the provision of the free Wi-Fi. For example, the neighborhood might include school children who are economically disadvantaged, and some people beyond the neighborhood could benefit from knowing that the children have Internet access to help them with their schoolwork. This would be an example of passive-use values.

Identification of the relevant study population for measurement of passive-use values can be challenging because there is unlikely to be a clear geographical boundary beyond which passive-use value falls to zero or to some minimum level. And the boundary issue can be of great importance, as in valuing a good or program of national interest but local implementation, such as preservation of a unique habitat or a loss caused by a major oil spill. The issue of whose benefits should be measured has been described by Smith (1993, p. 21) as "probably more important to the value attributed to environmental amenities as assets than any changes that might arise from refining our estimates of per unit values." Empirical studies of goods with large passive-use value components (Loomis 2000; Sutherland and Walsh 1985) have verified that only a small percentage of the aggregate economic benefit accrues to individuals in the immediate jurisdiction where the good is located. Unfortunately, no simple rule of thumb exists for defining the study population when measuring passive-use values. Specification of the study population requires thoughtful consideration of who will benefit from provision of the good and who might realistically be asked or required to pay for the good.

3.2.1.2 Sample Frame

The sample frame is the list from which sample units (i.e., the individuals who will be asked to complete the survey) are chosen. Ideally the sample frame perfectly matches the study population, but that is rarely the case. Strictly speaking, if the sample frame does not perfectly match the study population, generalization from the survey sample should be made only to the sample frame, not to the study population. In practice, such generalizations are made when the sample frame and study population are well, but not perfectly, matched. When the sample frame does not perfectly match the study population, coverage error occurs because not all members of the population have a known nonzero chance of being asked to complete the survey (Figs. 3.2 and 3.3).

"Undercoverage" is the part of the study population that is not included in the sample frame. For example, a survey of property owners, such as those with properties in the neighborhood where the Wi-Fi service would be provided in the previous example, would miss renters. Undercoverage will also occur if the survey is administered in a manner that excludes some individuals in the population. For example, a survey with an email letter of invitation would exclude individuals who

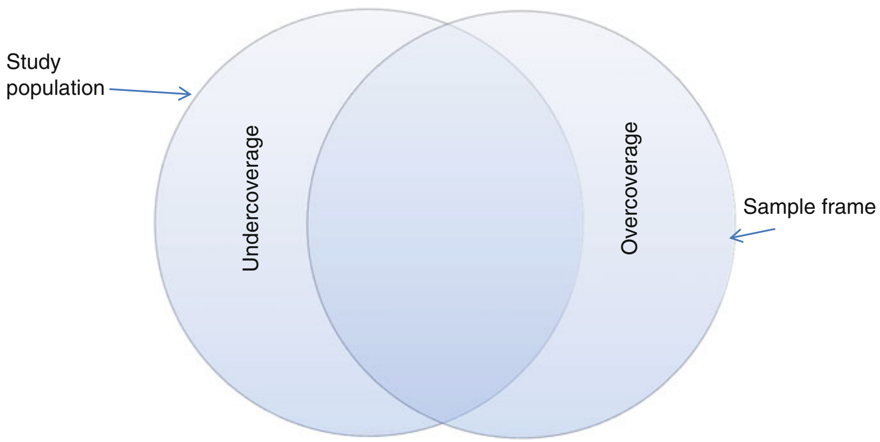


Fig. 3.3 Study population and sample frame

do not have an email address. Likewise, overcoverage occurs if the sample frame includes some individuals who are not in the study population, such as individuals who recently moved out of the study area or are deceased. Figure 3.3 illustrates the relationship between the study population and the sample frame.

Finding a single list of individuals or households in the study population with the requisite contact information might not always be possible. It is common for a sample frame to be developed from multiple sources, such as voter registration lists and tax assessor databases. For the Wi-Fi service case mentioned, if the researcher wants to measure the benefits to all residents in the neighborhood, she might consider a sample frame such as a property tax assessor database. Because this sample frame would fail to include individuals who are renters, she might augment the property tax assessor database with voter registration information. However, some residents might not be registered to vote, and the tax assessor database could include investors/landlords who are not residents. In such cases, purchasing a list of residents from a survey research firm might be a good option. Although coverage of the study population using such lists might not be perfect, it is often very good.

3.2.1.3 Drawing the Sample

After a sample frame has been developed, a sample can be chosen. There are two issues related to choosing a sample: how to choose sample units and how many sample units to choose. As mentioned previously, with probability sampling every unit in the study population has a known nonzero probability of being chosen; it is this fact that allows inferences to be made from the sample to the broader study population (or sample frame if the frame does not adequately represent the study population).

The most straightforward probability sampling technique is simple random sampling. With a simple random sample, every sample unit has an equal probability of being chosen, and it is not necessary to weight the final data (unless the data need to be weighted to correct for sample nonresponse).

However, many situations call for a more involved sampling technique than simple random sampling. For example, if the study population is geographically dispersed, a simple random sample can result in sample units that are also geographically dispersed. Geographic dispersion could be problematic for a survey that requires an interviewer to travel to the respondent to administer the questionnaire. An alternative in this case is cluster sampling, which can be used to minimize travel costs between on-site interviews. The general idea is that respondents are chosen in groups or clusters where the clusters are similar to each other, but the individuals within the clusters are representative of the population as a whole. For example, a simple cluster sample could define blocks in a city as clusters. First, blocks are selected and then potential survey participants within each of the blocks are chosen. Cluster sampling can be quite complicated, so consulting a survey statistician about sampling procedures is a good idea.

A third approach to random sampling, called "stratified random sampling," divides the sample frame into nonoverlapping subpopulations (strata) based on some measure available for the entire sample frame. Stratified sampling differs from cluster sampling in that the strata differ with respect to the measure (e.g., income, education) used to stratify the sample but the individuals within a stratum are similar to each other.

Within each stratum, a variety of methods can be used to select the sample units. Stratified random sampling is primarily used in three situations. First, it can be used to ensure adequate sample size within strata for separate analysis. This is important when individual strata are of particular interest. For example, consider a nonmarket valuation study with the primary objective of estimating mean willingness to pay for improved access to a recreational area and a secondary goal of comparing the mean willingness to pay between urban and rural populations. In this situation, a simple random sample might not provide an adequate number of rural respondents to allow for detection of a statistical difference. The sample could be stratified into rural and urban strata, with the rural stratum more heavily sampled to ensure an adequate number of respondents for analysis. The second situation in which stratified sampling is used is when the variance on a measure of interest is not equal among the strata. In the rural-urban example above, there could be higher variance of willingness to pay for the urban population relative to the rural population. In this case, more of the sample can be drawn from the stratum with the larger variance in order to increase the overall sample efficiency. The third situation in which a stratified sample might be used is when the cost of obtaining a survey response differs by strata. An example of this would be implementation of an online survey where possible, with a phone survey for individuals who do not have access to the Internet. Survey responses from the phone stratum will be more costly than those from the online survey stratum.

Beyond stratified and cluster sampling, there are many kinds of complex multistage random sampling approaches; see Groves et al. (2009) for a good explanation of these.

In addition to deciding how to choose sample units from a sample frame, the researcher must decide how many units to choose. Two important considerations in choosing the sample size are sampling error and statistical power. Sampling error (Fig. 3.2) is the deviation that arises because we observe statistics (e.g., mean willingness to pay or average income) for the sample rather than for the entire sample frame. Because the value of the population statistic is not known (if it were known, it would not be necessary to do the survey), the sampling error is approximated using a probabilistic approach that is a function of the sizes of the sample and the associated population. For small study populations, a relatively large proportion of the population is needed in the sample to maintain a given level of sampling error. For example, for a study population of size 1,000, approximately 200-300 observations (completed surveys) are required for ±5% sampling error. The sensitivity of the sampling error to sample size decreases as the population increases in size. Whether the size of the study population is 10,000 or 100 million, a sample size of approximately 380 is needed to obtain ±5% sampling error, which is not much more than the sample size needed for the previous example with a study population of 1,000 (see Dillman et al., 2009, p. 57, for a table of completed sample sizes for various population sizes and different margins of error).

The second consideration when choosing the sample size is statistical power. After the survey is complete, the researcher will often conduct statistical analyses in which a null hypothesis (H_0) and an alternative hypothesis (H_1) are specified. Statistical theory identifies two types of error that can occur when conducting hypothesis tests. Type 1 error occurs if H_0 is rejected when it is true. Type II error occurs if H_0 is not rejected when H_1 is true. The power function $(P(\theta))$ is defined as the probability of rejecting the null hypothesis when it is false (thus, it is 1 minus the probability of a type II error):

$$P(\theta) = 1 - \beta(\theta),$$

where θ is the parameter of interest and $\beta(\theta)$ is the probability of type II error. In other words, the power of a test is low when the probability of type II error is high. The details of estimating a power function can be found in most introductory statistics textbooks. In addition, Mitchell and Carson (1989, Appendix C) provided a nice exposition on estimating the power function for testing the difference between two means. One key point is that as the overall sample size increases, the power function increases. If statistical testing involves subsamples of the survey data, the number of observations in each of the subsamples must be considered. Of course the completed sample size is usually much smaller than the initial sample, so the initial sample should be selected with consideration of the expected response rate to provide an adequate number of final observations.

Finally, the sample size is also related to the survey mode (Sect. 3.2.3). With a fixed budget, choosing a more expensive mode of administration, such as an interviewer-administered survey, implies a smaller sample size than is possible with a relatively less expensive mode, such as a mail survey.

3.2.1.4 Nonprobability Sampling

There are situations in which inference to a broader population is not important. As described in Chap. 10, economic experiments often use nonprobability sampling to recruit participants. Likewise, online panel surveys (Sect. 3.2.3.1) typically use nonprobability samples. The three commonly used nonprobability sampling techniques are convenience, quota, and purposive samples. An example of a convenience sample would be recruiting students from a class to participate in a survey. In this situation, the survey respondents are recruited because it is easy, not because they have particular characteristics related to the research. On the other hand, with a purposive sample, individuals are intentionally recruited because they have characteristics related to the study that could be difficult to find through screening general populations. For example, choice experiments (Chap. 5) are often used to value the attributes of new food products. If a researcher would like to administer a choice experiment survey about lematoes, a fruit that is a cross of a tomato and a lemon, a purposive sample might include individuals in the grocery store who purchase tomatoes. Quota sampling is an approach where the individuals are recruited up to a predetermined number of respondents with particular characteristics. For example, it could be determined that the researcher wants 100 Nordic skiers and 100 snowmobilers to complete a survey on winter recreation in a state park.

There are many permutations to each of the three nonprobability sampling approaches described here. Strictly speaking, generalization beyond a nonprobability sample should not be done. However, as a practical matter, nonprobability sampling has become more common as response rates to probability samples have been declining. Furthermore, adjustments are sometimes made to data from nonprobability samples post hoc in an effort to draw inference to a population beyond the sample.[1] One limitation of post hoc adjustments is that the survey population measures for which data are available might not be strongly related to the nonmarket values of interest. For example, adjusting a sample based on U.S. Census data measures of education and income to make inference to the population from a sample about mean willingness-to-pay values for open space might not be appropriate if willingness to pay for open space is not related to income or education. It could be the case that willingness to pay for open space is strongly correlated with individuals' perspectives on the appropriateness of government-provided public goods. However, data are not likely available for the study population on such attitudes. The bottom line is that the researcher needs to be very clear about the nature of the sample and the appropriateness of drawing inferences from it.

3.2.2 Designing a Questionnaire

The previous section described approaches to sampling or rules for identifying potential survey respondents. Decisions about sampling and decisions about survey design are intertwined because both are affected by the same constraints (e.g., time, budget, available labor). For the sake of exposition, the survey design process is described before the mode of administration. In reality, many decisions related to questionnaire design and survey administration are made simultaneously. Regardless of how a survey is administered, the design process for the questionnaire is largely the same.

The design of the survey materials, including the questionnaire, recruitment materials, reminders, and any other items developed to communicate with potential participants, can affect whether a respondent completes and returns the

[1]Post hoc adjustments are weights applied to the data so that the sample has a distribution on key variables (usually demographic variables) that is similar to the population of interest. Groves et al. (2009) described approaches for weighting data. Post hoc adjustments are also made to data collected using probability sampling to correct for nonresponse error.

questionnaire. Fortunately, the researcher has control over the appearance of self-administered questionnaires, the tone and professionalism of interviewers, and the description of why an individual should complete the survey. The information in Sects. 3.2.2.1-3.2.2.4 highlight information relevant to all survey modes of administration described in subsequent sections. For more details on how to design and administer a survey, consult the text by Dillman et al. (2009), which is considered the survey research bible.

3.2.2.1 Identify Desired Measures

The first step in developing the questionnaire is to identify what it is that the researcher would like to measure in the survey. For example, the researcher may want to consider what data are needed to estimate the nonmarket valuation model, and test hypotheses of interest. The researcher will usually have a good sense of the goals of a study but could struggle with identifying the specific measures needed to address those study goals. After the researcher has identified the desired measures, he or she will need to develop questions to quantify the desired measures. For example, a researcher designing a contingent-valuation survey to measure willingness to pay for improved fishing conditions might also want to measure respondents' angling experiences. However, simply asking the question, "Are you an experienced angler?" will not provide useful information if some anglers consider skill as a measure of experience while others consider the number of times fishing as a measure of experience. The researcher might want to develop questions about the specific dimensions of angling experience, such as the number of years the respondent has been fishing, how often the respondent goes fishing, and the type of gear the respondent uses when fishing. It is often useful to review studies on similar topics to facilitate development of appropriate questions to capture desired measures.

3.2.2.2 The Questionnaire

When designing the questionnaire, the researcher wants to minimize the measurement error (Fig. 3.2) that arises if the responses to a survey question do not accurately reflect the underlying desired measure. Poor design of the survey questions causes measurement error. The questionnaire should be clearly written using common vocabulary. Focus groups and other qualitative approaches described in Chapter 8 of Groves et al. (2009) can be used to investigate issues such as appropriate language for the respondent population and baseline knowledge of the survey topic. How questions are phrased will vary with the mode of administration. Interviewer-administered questionnaires often pose questions in the spoken language, while self-administered surveys are usually written to be grammatically correct.

Contingent-valuation and choice-experiment surveys (Chaps. 4 and 5, respectively) often include a detailed description of the good or the program being valued. The researcher should make sure the description is accurate and consistent with scientific knowledge. However, it is essential that the vocabulary be appropriate for the survey respondents. For example, if the questionnaire includes a description of a program to eradicate white pine blister rust from high elevation ecosystems, the researcher cannot assume that the respondents know what white pine blister rust or high elevation ecosystems are. Such concepts need to be defined in the description of the program.

Three important issues related to the design of questions are discussed below. It might also be useful to read books specifically on survey design, such as Marsden and Wright (2010), Groves et al. (2009), Dillman et al. (2009).

One consideration related to question design is whether to ask the question using an open- or closed-ended format. Open-ended questions do not include response categories. For example, the question "What do you like most about visiting Yellowstone National Park?" suggests that the respondent is to answer the question in his own words. Responses to open-ended questions can be categorized subsequently, but doing so is subjective and time consuming. Nonmarket valuation surveys usually feature closed-ended questions, which are less burdensome to respondents than open-ended ones, and which are more readily incorporated into statistical analyses. Developing response categories for closed-ended questions takes careful consideration of the range of possible answers. In developing response categories, it can help to use open-ended questions in focus groups or one-on-one interviews (see Sect. 3.2.2.4).

Another issue is confusing question design. One type of confusing survey question is the double-barreled question. An example is asking respondents to indicate their level of agreement with the statement, "I am concerned about the environmental effects and cost of suppressing wildfires." A respondent who is concerned about the environmental effects of wildfires but not the cost of suppressing them will have a difficult time responding to this statement.

Survey respondents could also misinterpret attitude statements that switch between positive and negative as the following statements do:

- National parks should be easily accessible by roads.
- National parks should not include designated wilderness areas.
- National parks should be managed to preserve native plant species.
- Motorized vehicle should not be allowed in national parks.

Respondents who are asked whether they agree or disagree with these statements might not notice that the statements switch between positive and negative. While it is important that such statements not appear to be biased toward a particular perspective, the statements should all be stated in a consistent manner.

In addition, the response categories should be consistent with the question asked. For example, the question, "Did you go fishing in the last 12 months?" should have "yes" and "no" as response categories. However, it is not uncommon to see this type of question followed with response categories such as "frequently," "occasionally," and "never"; one would want to avoid this type of inconsistency between the question and the response categories. Further, if the reference time frame changes between questions, it should be pointed out to the survey respondent. For example, if one set of questions in a fishing survey asks about expenditures in the last 12 months and another set of questions refers to expenditures on the most recent trip, the change in reference should be explicitly noted.

How to elicit recall data is another question design issue. Frequency questions, such as those about the number of doctor visits in the last 12 months or the number of visits to a particular recreational area, are often posed in nonmarket surveys. Frequency questions can be difficult for some respondents. Schwarz (1990) summarized the cognitive processes that respondents go through when responding to behavioral frequency questions:

> First, respondents need to understand what the question refers to, and which behavior they are supposed to report. Second, they have to recall or reconstruct relevant instances of this behavior from memory. Third, if the question specifies a reference period, they must determine if these instances occurred during this reference period or not. Similarly, if the question refers to their "usual" behavior, respondents have to determine if the recalled or reconstructed instances are reasonably representative or if they reflect a deviation from their usual behavior. Fourth, as an alternative to recalling or reconstructing instances of the behavior, respondents may rely on their general knowledge, or other salient information that may bear on their task, to infer an answer. Finally, respondents have to provide their report to the researcher. (p. 99)

Respondents might be better able to recall frequency of events if they are provided with cues that make the questions more specific. For example, instead of asking how many times a person went fishing, asking the number of times a person went fishing for specific species could reduce the cognitive burden on the respondent. Likewise when asking about expenditures for travel to a particular recreation destination, it is helpful to ask about particular expenditures (e.g., fuel, food, equipment) rather than total trip expenditures. Given the heavy dependence of some nonmarket valuation techniques—such as the travel cost method—on recall data, the researcher needs to be sensitive to potential data-collection problems and to implement techniques to mitigate such problems. Chapter 6 has more specific suggestions for eliciting trip recall information.

3.2.2.3 Question Order and Format

After developing the questions, the order and format of the questions need consideration. Again, decisions about question order and format are closely related to whether the questionnaire is interview-administered or self-administered. Sudman and Bradburn (1982) suggested starting a questionnaire with easy, nonthreatening questions. Because contingent-valuation and choice-experiment questionnaires can require substantial descriptions of the goods or attributes of the goods, some practitioners break up these large descriptions by interspersing relevant questions about the provided information. Dillman et al. (2009) suggested putting objectionable questions, such as the respondent's income, at the end of the questionnaire because respondents are more likely to complete the entire questionnaire if they have already spent five or 10 min working on it. Asking all demographic questions at the end of the questionnaire is common practice.

It is important to realize that the order of the questions can influence responses. This phenomenon could be of particular concern for stated preference surveys that ask the respondent to value multiple goods. While the researcher can tell the survey respondent to value each good independently, individuals might not be willing or able to do so. Indeed there is much evidence of order effects associated with contingent-valuation questions (Powe and Bateman 2003). However, Clark and Friesen (2008) also found order effects associated with actual purchase decisions of familiar private goods and concluded that order effects could be associated with general preferences rather than exclusive to stated preferences. Such effects are difficult to mitigate, so it is often recommended that researchers randomize the order of questions that are susceptible to order effects.

3.2.2.4 Testing and Refining the Questionnaire

The final version of a questionnaire often bears little resemblance to the first draft. Numerous revisions can be required before arriving at a well-designed questionnaire, one capable of engaging the respondents and providing meaningful data. Qualitative methods such as focus groups and one-on-one interviews are commonly used to refine and develop the questionnaire. Because nonmarket valuation surveys can involve topics with which respondents are not familiar, understanding the respondents' baseline level of knowledge is essential to writing the questions at the appropriate level. Qualitative methods are also used to understand how respondents think and talk about a particular topic. The vocabulary of a highly trained researcher and a layperson can be quite different. The questionnaire should avoid jargon and technical language that is not familiar to respondents.

After the questionnaire has been drafted, it should be reviewed by experienced nonmarket valuation practitioners to assure that the appropriate measures are included and the design is conducive to a good response rate as well as to producing quality data. In addition, the questionnaire and other survey materials should be tested or reviewed by potential respondents using qualitative research methods. While such methods are generally considered helpful for developing and refining survey materials, there is no agreement about standard procedures. The most common qualitative approaches used in survey development—focus groups and one-on-one interviews are described below.

A focus group is a moderated meeting with a small group of individuals to discuss the survey topic as well as drafts of the questionnaire and other survey materials. The goal is to develop survey materials and identify potential problems. The discussion is led by a moderator who has an agenda that identifies the topics to be covered within the limit of the meeting time (usually two hours or less).

However, the moderator should allow for some unstructured conversation among the participants. The balance between covering agenda items and allowing a free flow of conversation can pose a challenge to the moderator.

Another challenge is dealing with the diverse personalities of the focus group participants. A participant with a strong personality can dominate the focus group if the moderator is not able to manage that participant and draw out comments from the more reserved participants. The researcher should carefully consider how the small-group dynamic can influence what issues are raised and how important participants really think those issues are. For example, participants are sometimes susceptible to jumping on the bandwagon regarding issues that they would not have considered otherwise.

Some project budgets will allow for focus groups conducted at professional focus group facilities by an experienced moderator. However, much can also be learned using less expensive focus groups conducted in a more informal setting such as a university classroom or local library. The important attributes of a successful focus group include development of an appropriate agenda, a somewhat diverse group of participants, and a moderator who is adequately trained to obtain the desired information from the participants. Focus groups are usually recorded (audio and/or video) so the moderator can concentrate on the group discussion without being concerned with taking notes. Afterward, the recordings are reviewed and summarized. Additional details about how to conduct a focus group can be found in Greenbaum (2000).

Another qualitative method used to develop and refine nonmarket survey materials is the one-on-one interview, also called a cognitive interview (Groves et al. 2009). A one-on-one interview includes an interviewer and a potential survey participant. Unlike focus groups, the intimate context of a one-on-one interview removes the possibility of a participant being influenced by anyone other than the interviewer. These interviews usually take one of two forms: either the survey participant is asked to complete the questionnaire and review accompanying materials, such as a cover letter, while verbalizing all of his or her thoughts, or the interviewer asks questions from a structured list and probes respondents about their answers. In the latter case, after the respondent has answered the survey questions, the interviewer might ask about motivations for responses, interpretation of a question, or assumptions made when answering the questions. Enough interviews should be conducted to allow for a sense of whether issues are particular to one individual or are more widespread.

3.2.3 Survey Mode

Tradeoffs encountered in choosing the best mode for administering the survey include survey administration costs, time constraints, sample coverage, and context issues. Given the rapid changes in communication technology, it is best to have the survey administration plan reviewed by an experienced survey researcher because there could be approaches not considered or unforeseen pitfalls of a mode under consideration.

In general, surveys can be administered in two ways: a respondent can complete a questionnaire (self-administered), or an interviewer can orally administer the questionnaire and record the responses (interviewer-administered). Of course there are hybrid options as well. For example, questions can be asked orally and responses made in written form or entered on a keypad. The next two sections summarize key considerations when choosing between a self-administered survey and an interviewer-administered survey.

3.2.3.1 Self-Administered Surveys

A self-administered survey can be carried out using paper or an electronic device. Paper surveys are often mailed to potential respondents, but they can also be handed out at a recreation site or other location where the respondents have gathered. Electronic surveys (e-surveys) can be completed online, as an attachment to or embedded in an email, or on a device provided by the survey sponsor. For example, tablet computers can be used to complete questionnaires on-site. Self-administered surveys usually require fewer resources to implement than interviewer-administered surveys and, in turn, are often less expensive. The procedures for implementing a self-administered survey are also less complicated than those of interviewer-administered survey. Furthermore, interviewer effects are avoided, and the questionnaire can be completed at the respondent's pace rather than the interviewer's.

The main drawback of self-administered surveys is that there are several potential sources of error. Nonresponse errors (Fig. 3.2) can arise if respondents examine the survey materials to see if they are interested in the topic and then decide not to participate. This can result in important differences between respondents and nonrespondents. In addition, completion of a self-administered questionnaire requires that respondents be competent readers of the language.

An additional potential drawback of a self-administered paper questionnaire is the loss of control over the order in which questions are answered. Respondents to a self-administered paper questionnaire can change their responses to earlier questions based on information presented later in the survey. It is not clear if this would be desirable or undesirable; however, it is an uncontrollable factor. Control over order can be maintained in a self-administered e-survey. Finally, if the survey is self-administered via the postal service, there can be a long time lag between when the survey packet is sent out and when the questionnaire is returned.

E-surveys are gaining in popularity and offer many advantages over both self-administered paper surveys and interviewer-administered surveys. The cost per completed e-survey can be lower than self-administered paper surveys or interviewer-administered surveys. In addition, the responses to an e-survey are entered electronically by the survey respondents, saving time and money. The downside to an e-survey is that it can be completed on different types of devices, such as laptop computer, a tablet computer, or a smart phone, and formatting for all the different devices can be tricky. Also, setting up an e-survey can require expertise in computer programming. However, there are programs available that can be used to set up a relatively straightforward e-survey. Lindhjem and Navrud (2011) reviewed stated-preference nonmarket studies that compare e-surveys to other modes of administration and also reported on their own contingent-valuation study that compared results of an in-person, interview-administered survey to an e-survey administered online. They found that the share of "don't knows," zeros, and protest responses to the willingness-to-pay question was very similar between modes, and equality of mean willingness-to-pay values could not be rejected.

Self-administered e-surveys are often administered to online panels made up of individuals who agree to complete multiple e-surveys. Numerous approaches are used to recruit panel participants. While some panelists could continue to participate indefinitely, attrition is an ongoing issue, so new panel members might need to be continuously recruited. Panel participants can refuse to complete any individual survey and are often compensated for their participation. Some, but not most, online panels are recruited using probability-based methods. In such cases, some recruits might not have Internet access, so it is provided by the agency recruiting the online panel.

When a researcher contracts with a company to implement a survey online, the company will usually agree to provide a specified number of completed surveys. The costs of panel online surveys are usually lower than those of an interviewer-administered survey and might be less than a mail survey. If the researcher wants to generalize the results of the survey to a broader population, the panel company might guarantee that respondent demographics match the population of interest, or the survey data could be weighted to match the demographic characteristics of the study population. However, using weights with a nonprobability sample might not be sufficient for making inference to a broader population. As discussed in Sect. 3.2.1.4, such adjustments can only be made using measures that are observable for the broader population (i.e., demographic characteristics). If the observable characteristics of the broader population are found to be unrelated to the nonmarket values in a statistically significant manner, weighting the panel data to draw inference about the nonmarket values of a population broader than the panel participants is called into question. In addition, a nonprobability online survey sample does not provide a framework, such as that provided in Fig. 3.2, for identifying sources of error (Tourangeau et al. 2013).

The advantage of a panel survey to the researcher is the guarantee of a given number of completed surveys from respondents with demographics that match a broader population. Further, the surveys are usually administered quickly, so the researcher has data in hand in a short period of time. However, not all panel companies are equivalent, so the researcher should investigate the methods of a particular company.

The American Association for Public Opinion Research provides recommendations regarding use of online panels (Baker et al. 2010). Its first recommendation is "Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values." However, the association also concludes, "There are times when a nonprobability online panel is an appropriate choice. Not all research is intended to produce precise estimates of population values, and so there could be survey purposes and topics where the generally lower cost and unique properties of Web data collection are an acceptable alternative to traditional probability-based methods." Finally, it makes clear that users of online panels need to make sure they understand the nature of the particular panel: "Users of online panels should understand that there are significant differences in the composition and practices of individual panels that can affect survey results. Researchers should choose the panels they use carefully" (p. 714).

At the request of the U.S. Environmental Protection Agency, the Wyoming Survey and Analysis Center recently compared results obtained using an online panel with those obtained from mail and phone surveys. The sample for the mail and phone surveys was based on a random-digit dialing list with reverse address look-up purchased by the Wyoming Survey and Analysis Center from a vendor. Knowledge Networks, the company that administered the online panel, used a very similar sample for recruitment into the panel. The details of the sampling are described in Grandjean et al. (2009), Taylor et al. (2009). The authors found that response rate was highest for the mail survey (30%), followed by the phone survey (16%), then the online panel survey (4%). The mail and phone survey response rates were computed using a comprehensive formula identified by the American Association for Public Opinion Research (Smith 2015) that accounts for not only the individuals known to be eligible (i.e., they were successfully contacted and met the criteria for participation) but also for the individuals who were never successfully contacted in initial phone efforts. The bottom line is that comprehensive estimates of response rates are low for all three modes, but they are particularly low for the online panel. In the online panel survey, many potential respondents were lost because they either did not agree to participate in the panel initially or they left the panel prior to the survey being implemented. However, the completion rate (i.e., number of completed surveys per number of individuals identified as potential participants) is highest for the phone survey (91%), followed by the online survey (77%) and the mail survey (75%). As is evident in this study, completion rates can be quite different from response rates.

Grandjean et al. (2009) also found some statistically significant differences across modes in the demographic characteristics of the respondents. For example, the online panel participants were less likely to belong to an environmental organization than the phone or mail respondents. The measure of interest in this comparison study was willingness to pay for cleaner air in national parks. The willingness-to-pay estimate based on the phone survey was higher than the willingness-to-pay estimate from the mail or online panel surveys. The authors speculated that this result is due to social desirability bias (described in the following section) associated with the phone survey and concluded that the online panel survey provided a willingness-to-pay estimate as accurate as the well-designed mail survey. However, without a thorough investigation of potential nonresponse bias, it is not clear how the willingness-to-pay estimates from the survey respondents compare to those of the broader population.

3.2.3.2 Interviewer-Administered Surveys

Interviewer-administered surveys have traditionally been conducted in-person or over the telephone. However, current technology also allows for an interviewer in a remote location to administer a survey via technology that transmits video, voice, or text over the Internet. Interviewers can administer surveys in respondents' homes, in convenient locations such as shopping malls or on the street, or in places where the participants of targeted activities tend to congregate. For example, recreation use surveys are often administered at the site of the recreational activity (e.g., a hiking trailhead, boat landing, or campsite).

The primary advantages of interviewer-administered surveys come from the control the interviewer has while administering the survey. The interviewer can control who completes the survey as well as the order of questions. In addition, the interviewer can provide clarification if the survey is complex. If a survey is administered in-person, a system can be developed for choosing participants that avoids the need for an initial sample frame with contact information. Such an approach is often used in countries where complete lists of individuals who reside in an area are unavailable. Finally, if an interviewer administers a survey, the respondent does not need to be literate.

An important downside of an interviewer-administered survey is that it tends to be more expensive than a self-administered survey due to interviewer costs. The interviewers need to be paid for training, conducting the interviews, and travel costs if the surveys are conducted in-person. Another major concern with interviewer-administered surveys is that the presence of an interviewer can affect survey responses (Singer and Presser 1989). In particular, social desirability bias—where the presence of the interviewer results in a more socially desirable response (e.g., a higher willingness to pay or higher number of trips to a recreation site reported) than would be elicited in a self-administered survey—is a concern for nonmarket valuation surveys.

3.2.4 Administering the Survey

If the questionnaire and other survey materials are finalized, the plan for drawing the sample completed, and the survey mode selected, the survey is almost ready for administration. The next two sections cover approaches for contacting potential survey respondents and for conducting a pilot test of the survey procedures before final implementation.

3.2.4.1 Contacting Potential Survey Respondents

Recruitment of potential survey respondents need not employ the same mode as the administration of the final survey. Potential survey respondents can be contacted via the postal service, email, social media sites, the telephone, or in-person (e.g., at their residences or a public location, such as a recreation site or a mall). The information in the sample, such as telephone numbers and mailing addresses, will dictate the options for contacting potential survey respondents. Sometimes the invitation to complete the survey and the subsequent administration of the survey are done together. For example, if the sample includes address information, a survey packet can be mailed with a cover letter inviting potential respondents to complete an enclosed paper questionnaire. As mentioned earlier, mixed-mode surveys are currently the norm. For example, a letter could be sent through the mail to invite a potential respondent to complete an online survey.

3.2.4.2 Pilot Survey

A pilot survey is a small-scale test of the draft survey materials and implementation procedures. The pilot survey serves several purposes. First, it allows for a test of the survey sample. Some samples can be problematic, but the researcher might not realize it until the survey is in the field and great expense has already been incurred. For example, I was involved in the design of a mail survey in which a nonprobability sample was developed using a list of registered voters.[2] Even though the list was current, during the pilot study we found that many of the addresses were not valid, and a substantial number of the surveys were undeliverable. In the end, a completely different sample was used for the final survey.

The pilot survey also provides information about the anticipated response rate, the number of follow-ups that will be needed, and the expected cost for the final survey. Third, the pilot study allows for a test of the survey implementation procedures. For contingent-valuation studies, the pilot study also offers an opportunity to learn about the distribution of willingness to pay. For example, if the researcher plans to use a dichotomous-choice contingent-valuation question in the final survey, an open-ended question can be used in the pilot study to guide assignment of the offer amounts for the final survey (see Chap. 4 for details of contingent-valuation surveys).

[2]The study was a split sample test of different decision rules in a contingent-valuation question. Because we did not need to draw inference to a larger population about the magnitude of the estimated contingent values, a nonprobability sample was used.

3.2.4.3 Survey Implementation

After the survey materials have been designed and the implementation procedures tested, the survey is ready to go into the field. Regardless of the mode of administration, final survey implementation tends to be hectic. A mail survey requires a substantial effort. The questionnaire and cover letters need to be printed, the postage and envelopes need to be purchased, and the survey packets need to be assembled for mailing. If a survey is going to be interviewer-administered, the interviewers must be trained in the survey administration procedures.

One approach to survey implementation is to offer multiple modes for completing the survey. For example, respondents can be offered a choice of completing an online or a mail questionnaire. The motivation for a mixed-mode survey is that some respondents might be more likely to respond to their preferred mode, resulting in a higher response rate than with a single-mode survey. As part of an annually administered survey, Olson et al. (2012) asked respondents in 2008 about their preferred mode for future surveys. They found that for Web-only and phone-only mode preferences, the response rate was significantly higher for the individuals who received their preferred mode compared to those who did not receive their preferred mode in the 2009 survey. Likewise, results of a meta-analysis suggest that mail surveys with a simultaneous Web option have lower response rates than a mail-only option (Medway and Fulton 2012). The bottom line is that intuition that mixed modes will increase response rates is not backed up with strong evidence.

A database should be set up with the information from the sample that allows for tracking respondents and nonrespondents. In addition to tracking whether an individual has responded, it is beneficial to track when an individual responded and when follow-up contacts were made. This information can be used for planning future surveys, and it allows for testing for differences between early and late respondents.

Because most nonmarket valuation surveys elicit sensitive information such as annual household income, it is important to assure respondents that their responses will be confidential and their names will not be associated with their responses. The standard procedure is to assign a unique identification number to each sample unit in the initial sample to track whether or not the person has responded. A unique identification code should also be used with online surveys to make sure that only those invited to complete the survey do so and that no one completes the survey more than once. The identification numbers, not the respondents' names or addresses, are included in the final data set. Likewise, survey interviewers have an ethical responsibility to keep all responses confidential.

Surveys conducted through universities usually require informed consent, which involves informing respondents that their participation is voluntary. The specific information provided to respondents about the terms of their participation and the procedures for getting permission from a university to do a survey vary among universities. The funding organization for the survey should be knowledgeable about both the procedures for obtaining permission to conduct a survey and the procedures for obtaining consent from the survey respondents.

3.2.4.4 Survey Responses

Ideally, all contacted individuals would complete the survey; however, only a portion of the contacted individuals will actually respond. The response rate to a survey is the number of completed surveys divided by the number of individuals who were invited to complete the survey. As long as the individuals who respond to the survey are not different from those who do not respond to the survey in a manner that is relevant to the study objectives, a response rate of less than 100% is of no concern, and nonresponse error is nil. However, a synthesis of the research on nonresponse (Groves 2006) suggests that nonresponse error is present in many surveys.

There is confusion in the literature about how nonresponse error is related to the response rate. It is often suggested that a higher response rate will decrease nonresponse error. For example, the guidelines for statistical surveys sponsored by U.S. government agencies suggest that an analysis of nonresponse error should be conducted for response rates lower than 80% (U.S. Office of Management and Budget 2006). Groves and Peytcheva (2008) conducted a meta-analysis of the effect of response rates on nonresponse bias and found that the response rate by itself was not a good predictor of nonresponse bias. In other words, nonresponse bias was found in some surveys with relatively high response rates and not found in some surveys with relatively low response rates. They did find that general population surveys, surveys sponsored by government agencies, and surveys of study populations that did not have previous knowledge about the study sponsor were associated with higher nonresponse bias. There is no hard rule about what is an acceptable response rate; rather, a reasonable response rate is one that provides enough observations for the desired analyses and a group of respondents with characteristics similar to the population to which inferences will be made. If possible, an analysis of nonresponse bias should be conducted.

3.2.5 Completed Surveys

If the surveys were completed electronically, a useable dataset will be provided automatically. In other cases, several additional steps are required. The first step is to develop a set of rules, called a codebook, for data entry. Then the data are entered and verified. Finally, data analyses can be conducted.

3.2.5.1 Codebook

Coding refers to the process of translating the text responses into numerical responses. For example, a question that elicits a yes or no response could be coded so that a “1” represents a yes response and a “0” represents a no response. A codebook is developed for the entire questionnaire; it specifies the variable names

Have you ever visited Yellowstone National Park? (Circle one number)

1 Yes

2 No

Fig. 3.4 Sample survey question

for each survey question. One option to make the coding process easier is to include the numerical codes next to the response categories on the questionnaire, as shown Fig. 3.4. If, however, the responses are recorded by filling in a bubble or putting an "X" in a box, a numerical code might need to be written on the survey prior to data entry. Even when responses have corresponding numbers, some judgment can be involved in coding the data if the respondent did not follow the directions. Therefore, in addition to simple rules, such as 1 = yes and 0 = no, the codebook specifies numerical codes related to skipping a question. Weisberg et al. (1996) described standard coding practices.

Coding open-ended questions is more complicated. One option is to transcribe entire responses to open-ended questions. Analyzing such data would require use of a qualitative method such as content analysis. Another approach is for the researcher to develop a workable number of response categories and code each open-ended response into the appropriate category.

3.2.5.2 Data Entry

If surveys are not administered electronically, the data must be entered into an electronic database. Improvements in database management software have made data entry fairly easy. The software can be set up to notify the data entry person when a number that is out of the relevant range has been entered or to skip questions that the respondent was supposed to skip but did not. These innovations minimize data entry errors. However, even when using a data entry software package, the entered data should be checked. One option is to randomly compare surveys responses and the electronic data file. Another option is to enter the data twice. During the second entry of the data, the computer software signals if an entry does not match the original data entry.

After the data have been entered, frequencies should be run on all variables and checked to make sure the response values are within the appropriate range for each question and that skip patterns have been properly followed. Since a unique identification number is entered in the dataset and written on the actual questionnaire, errors in the data can be checked by going back to the original surveys. This step is referred to as data cleaning.

If the survey was conducted with the intent of generalizing to a broader population, the socioeconomic variables from the survey respondents should be compared to those of the broader population. If the survey respondents differ from the broader population to which survey results are to be generalized, sample weights can be applied. Quite a few articles describe how to weight respondent populations (see Whitehead et al. 1993, 1994; Dalecki et al. 1993; Mattsson and Li 1994; Whitehead 1991).

3.2.5.3 Documentation

Regardless of the study objectives, all studies should include documentation of the procedures used to collect the data as well as summary statistics on all measures elicited in the survey. Proper documentation will make the data more accessible. Chapter 11 describes the benefit transfer approach, where results from an original study are used in a different context. For an original study to be amenable to benefit transfer, the original study must be appropriately documented. Complete documentation will also assist reviewers of the original study. Loomis and Rosenberger (2006) provided suggestions for reporting of study details to improve the quality of benefit transfer efforts. The best way to document data collection procedures is to do it as the study is being conducted. The entire study—from the initial study plan to the final summary statistics—should be documented in one location. Table 3.1 outlines the information about the study that should be documented. In addition, data should be archived. Oftentimes the funding institution provides guidelines on how data should be archived.

3.3 Secondary Data

Nonmarket values can often be estimated using secondary (i.e., existing) data. Secondary data might or might not have been collected with the original intent of estimating nonmarket values. Although use of secondary data can save substantial expense and effort, secondary data should only be used if the researcher is confident about the quality. In the U.S. many large national surveys, such as the U.S. Decennial Census, are well-designed and collect data that could be used for nonmarket valuation. The U.S. Census Bureau administers many other useful demographic and economic surveys, such as the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. Most of the data collected by the Census Bureau can be downloaded from its website (www.census.gov). Likewise, Eurostat, the statistical office of the European Union, provides statistics and data from European countries (www.epp.eurostat.ec.europa.eu).

An incredible amount of data from cell phones, the Internet, Global Positioning Systems, and other sources is continually being created. These so-called “big data” sources are a relatively new development and have not been used extensively in nonmarket valuation. However, it seems likely that these new data sources will soon provide important valuation opportunities. One can imagine the travel cost models described in Chap. 6 using data provided by a fishing app (application

Table 3.1 Guidelines for documenting a study

	Item to document	Information to include
Step 1	Study plan	Date Initial budget Study proposal
Step 2	Focus groups/one-on-one interviews	Dates Location Costs (amount paid to participants, food, facility rental, etc.) Focus group moderator/interviewer Number of participants (do not include names or addresses of participants) Handouts Agenda Video and/or audio recordings Interview or focus group summaries
Step 3	Sample	Description of study population Description of sample frame Procedures for selecting sample units Cost of sample Number of sample units Data included with sample
Step 4	Final survey	Text of letter or script used to invite potential respondents Questionnaires Dates of all survey contacts Number of respondents to each wave of contact
Step 5	Data	Codebook Final sample size Frequencies (including missing values) for each measure in the survey Means/medians for continuous measures Data file in format useable for data analysis software

software) or from information about visitation to rock climbing areas from GPS units. Cukier and Mayer-Schoenberger (2013) provided an overview of what big data sources are and how they are changing the way we approach data. Of course, while the allure of huge samples and lots of information make big data an exciting proposition for nonmarket valuation, the researcher should weigh these benefits against the structure and control available with an original survey.

Another secondary data source often used in nonmarket valuation is confidential data. For example, the information used to estimate damage costs described in Chap. 8 often includes hospital and/or urgent care admissions data. Use of such data requires strict adherence to procedures developed to maintain confidentiality. The provider of such data might require that no results be reported for small subsamples that could potentially allow for identification of individuals affected by the disease or health symptom of interest. The application process for obtaining confidential data usually specifies how the data can be stored, used, and reported.

A consideration when using any kind of secondary data is whether the data contain the requisite variables for estimating the desired nonmarket value. It is not only an issue of having the variables or not—the variables must be in the appropriate units. If, for example, one wants to estimate a travel cost model using some expenditures that are in annual terms and other expenditures that are reported on a per-trip basis, the expenditures must be converted so that both are on either a per-trip basis or an annual basis. Another consideration is whether the data are representative of the appropriate population. The key sampling issues that need to be considered when collecting primary data, such as defining the relevant study population and making sure the final data are representative of that population, should also be considered when assessing the appropriateness of secondary data.

Geospatial data are often used to augment nonmarket valuation data. One such source is Landsat data (http://landsat.gsfc.nasa.gov/data/), which uses satellite imagery of the earth to monitor changes in land use, water quality, encroachment of invasive species, and much more. These data are maintained by the U.S. Geological Survey and are available for free (http://landsat.usgs.gov/). Researchers can use these data in nonmarket valuation models as an objective measure of the change in the quality or quantity of an environmental amenity (or disamenity). When augmenting a nonmarket valuation model with geospatial data, the researcher must ensure that the scales of the different data sources are similar.

3.4 Conclusion

Many individuals come to nonmarket valuation with a background in economics that facilitates understanding of the associated theory and estimation methods. However, few newcomers to nonmarket valuation arrive with survey research experience. The goal of this chapter was to provide guidance on how to collect primary data or assess the quality of secondary data to support estimation of valid nonmarket values.

A survey effort starts with identifying the relevant study population. This straightforward-sounding task involves many considerations, such as who will benefit from provision of the nonmarket good and who will pay for it. The researcher must also consider how potential respondents will be contacted and how the survey will be administered. Rapid changes in communication technology are opening new options for administering surveys while also complicating many aspects of traditional survey research methods.

Perhaps the most difficult task in developing a nonmarket valuation survey is moving from notions of what the researcher would like to measure in a survey to the articulation of actual survey questions. This chapter provided general guidelines to help nonmarket valuation newcomers through the survey process.

Now that the general economic perspective on nonmarket valuation has been introduced in Chap. 1, the conceptual framework presented in Chap. 2, and nonmarket data collection methodology described in this chapter, the reader is prepared for the next eight chapters that describe specific nonmarket valuation techniques. The last chapter in this book focuses on reliability and validity of nonmarket valuation. As that chapter discusses, high quality data alone will not guarantee the reliability or validity of nonmarket values; however, quality data are a prerequisite for reliable and valid measures.

References

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J., Lavrakas, P. J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R. K. & Zahs, D. (2010). AAPOR report on online panels. Public Opinion Quarterly, 74, 711-781.

Clark, J. & Friesen, L. (2008). The causes of order effects in contingent valuation surveys: An experimental investigation. Journal of Environmental Economics and Management, 56, 195-206.

Cukier, K. N. & Mayer-Schoenberger, V. (2013). The rise of big data: How it's changing the way we think about the world. Foreign Affairs, 92, 28-40.

Dalecki, M. G., Whitehead, J. C. & Blomquist, G. C. (1993). Sample non-response bias and aggregate benefits in contingent valuation: An examination of early, late and non-respondents. Journal of Environmental Management, 38, 133-143.

Dillman, D. A., Smyth, J. D. & Christian, L. M. (2009). Internet, mail and mixed-mode surveys: The tailored design method (3rd ed.). New York: John Wiley & Sons.

Grandjean, B. D., Nelson, N. M. & Taylor, P. A. (2009). Comparing an Internet panel survey to mail and phone surveys on willingness to pay for environmental quality: A national mode test. Paper presented at the 64th annual conference of The American Association for Public Opinion Research, May 14-17. Hollywood, FL.

Greenbaum, T. L. (2000). Moderating focus groups: A practical guide for group facilitation. Thousand Oaks, CA: Sage Publications.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70, 646-675.

Groves, R. M. & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly, 72, 167-189.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). Survey methodology (2nd ed.). New York: John Wiley & Sons.

Lindhjem, H. & Navrud, S. (2011). Using internet in stated preference surveys: A review and comparison of survey modes. International Review of Environmental and Resource Economics, 5, 309-351.

Loomis, J. B. (2000). Vertically summing public good demand curves: An empirical comparison of economic versus political jurisdictions. Land Economics, 76, 312-321.

Loomis, J. B. & Rosenberger, R. S. (2006). Reducing barriers in future benefit transfers: Needed improvements in primary study design and reporting. Ecological Economics, 60, 343-350.

Marsden, P. V. & Wright, J. D. (Eds.). (2010). Handbook of survey research (2nd ed.). Bingley, United Kingdom: Emerald Group.

Mattsson, L. & Li, C. Z. (1994). Sample nonresponse in a mail contingent valuation survey: An empirical test of the effect on value inference. Journal of Leisure Research, 26, 182-188.

Medway, R. L. & Fulton, J. (2012). When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates. Public Opinion Quarterly, 76, 733-746.

Mitchell, R. C. & Carson, R. T. (1989). Using surveys to value public goods: The contingent valuation method. Washington, DC: Resources for the Future.

Olson, K., Smyth, J. D., & Wood, H. M. (2012). Does giving people their preferred survey mode actually increase survey participation rates? An experimental examination. Public Opinion Quarterly, 76, 611-635.

Powe, N. A. & Bateman, I. J. (2003). Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs. Ecological Economics, 45, 255-270.

Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), Research methods in personality and social psychology (pp. 98-119). Beverly Hills, CA: Sage.

Singer, E. & Presser, S. (Eds.). (1989). Survey research methods: A reader. Chicago: University of Chicago Press.

Smith, V. K. (1993). Nonmarket valuation of environmental resources: An interpretive appraisal. Land Economics, 69, 1-26.

Smith, T. W. (Ed.). (2015). Standard definitions: Final dispositions of case codes and outcome rates for surveys (8th ed.). American Association for Public Opinion Research. Oakbrook Terrace, IL: AAPOR. www.aapor.org/AAPORKentico/AAPOR_Main/media/publications/Standard-Definitions2015_8theditionwithchanges_April2015_logo.pdf.

Sudman, S. & Bradburn, N. M. (1982). Asking questions. San Francisco: Jossey-Bass.

Sutherland, R. J. & Walsh, R. G. (1985). Effect of distance on the preservation value of water quality. Land Economics, 61, 281-291.

Taylor, P. A., Nelson, N. M., Grandjean, B. D., Anatchkova, B. & Aadland, D. (2009). Mode effects and other potential biases in panel-based internet surveys: Final report. Wyoming Survey & Analysis Center. WYSAC Technical Report No. SRC-905. Laramie: University of Wyoming. http://yosemite.epa.gov/ee/epa/eerm.nsf/Author/A62D95F235503D03852575A800674D75.

Tourangeau, R., Conrad, F. G. & Couper, M. P. (2013). The science of web surveys. New York: Oxford University Press.

U.S. Office of Management and Budget. (2006). Standards and guidelines for statistical surveys. Washington, DC: OMB.

Weisberg, H. F., Krosnick, J. A. & Bowen, B. D. (1996). An introduction to survey research, polling, and data analysis (3rd ed.). Thousand Oaks, CA: Sage Publications.

Whitehead, J. C. (1991). Environmental interest group behavior and self-selection bias in contingent valuation mail surveys. Growth and Change, 22, 10-21.

Whitehead, J. C., Groothuis, P. A. & Blomquist, G. C. (1993). Testing for non-response and sample selection bias in contingent valuation. Economics Letters, 41, 215-220.

Whitehead, J. C., Groothuis, P. A., Hoban, T. J. & Clifford, W. B. (1994). Sample bias in contingent valuation: A comparison of the correction methods. Leisure Sciences, 16, 249-258.

Chapter 4
Contingent Valuation in Practice

Kevin J. Boyle

Abstract Contingent Valuation in Practice takes the reader through each of the basic steps in the design and administration of a contingent-valuation study. The text is written for the novice conducting their first study, decision makers who might want to better understand the method, and experienced researchers. Rich citations to the literature are provided to help a novice reader gain a deeper understanding of the supporting literature and to assist experienced researchers in developing the context for research to enhance contingent valuation in practice. The novice and users of value estimates can rely on the content of this chapter to understand best current practices and researchers can use the content to identify research that addresses uncertainties in the design and conduct of contingent valuation studies. Following the steps outlined in the chapter provides a basis to establish that an empirical study satisfies conditions of content validity.

Keywords Contingent valuation · Stated preference · Nonmarket valuation · Survey research · Hicksian suplus

Contingent valuation is a stated preference method and a survey-based approach to nonmarket valuation. A contingent-valuation question carefully describes a stylized market to elicit information on the maximum a person would pay (or accept) for a good or service when market data are not available. While controversial, as will be discussed below, contingent valuation and choice experiments (Chap. 5)—a close cousin in the stated preference family of valuation methods—are arguably the most commonly used nonmarket valuation methods.

An early contingent-valuation study was conducted by Davis (1963) to estimate the value of big game hunting in Maine. About a decade later, Cicchetti and Smith (1973) and Hammack and Brown (1974) applied contingent valuation to wilderness recreation and waterfowl hunting. Simultaneously, an application to valuing visibility in the Four Corners region of the Southwestern United States represented a turning point after

K.J. Boyle (✉)
Virginia Tech, Blacksburg, VA, USA
e-mail: kjboyle@vt.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_4

which contingent valuation gained recognition as a methodology for estimating values for public goods (Randall et al. 1974). The types of environmental applications have continually expanded, and studies have spread to health, cultural, and other types of applications when market data are not available to support decision-making.

Results from early applications of contingent valuation met with skepticism and criticism. One of the more notorious comments was expressed by Scott (1965), who referred to contingent valuation as a "short cut" and concluded, "ask a hypothetical question and you get a hypothetical answer" (p. 37). The contingent-valuation question is hypothetical in the sense that people do not actually make a monetary payment. Some of this criticism was deflected by Bishop and Heberlein's (1979) landmark validity study in which they compared welfare estimates for goose hunting from actual cash transactions, contingent valuation, and travel cost (Chap. 6). This study showed that the contingent-valuation estimate of willingness to pay (WTP) was of similar magnitude to estimates of WTP provided by a travel cost model and the cash transactions. Comparison of the contingent values to the travel cost values is a test of convergent validity, and comparison with cash transactions is a test of criterion validity (Carmines and Zeller 1979).

A workshop sponsored by the U.S. Environmental Protection Agency was the first attempt to synthesize what was known about contingent valuation (Cummings et al. 1986). The notable outcome of this "state-of-the-art assessment" was a set of reference operating conditions for conducting a credible contingent-valuation study. The reference operating conditions motivated much research to evaluate the validity of contingent valuation and to probe the limits of the types of applications where contingent valuation could provide credible welfare estimates.

Perhaps the most substantive contribution to the contingent-valuation literature was made with a book by Mitchell and Carson (1989) that presented detailed discussion on designing a contingent-valuation study. The book provided the broad overview that novices require when conducting their first contingent-valuation study, as well as prescriptive recommendations that set off a whole new wave of validity research. Mitchell and Carson fundamentally shifted the research focus to one that considered the details of study design. Validity, rather than being a global, all-or-nothing criterion, was now viewed as a function of specific aspects of study design; specific design elements were found to create or reduce bias in estimates of central tendency and increase or decrease the dispersion around estimates of central tendency.

Through the first 25 or more years of the use of contingent valuation, critiques of this methodology seemed to ebb and flow without a specific focal point of attack. This all changed when contingent-valuation estimates began to be used in legal cases as the basis of damage payments by parties responsible for large-scale pollution under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) of 1980 (Ward and Duffield 1992). The controversy became particularly heated after the settlement of the Natural Resources Damage claim for the Exxon Valdez oil spill. Exxon supported the publication of a book that critiqued the fundamental premises of contingent valuation (Hausman 1993), and the National Oceanic and Atmospheric Administration (NOAA) responded with a blue ribbon panel to evaluate the credibility of using contingent valuation to estimate passive-use values (Arrow et al. 1993). Passive-use values are those values held by individuals for the item being valued itself and not for any personal use of the item (see Chap. 2, Eq. (2.26)) and are also commonly referred to as nonuse values.[1]

The NOAA panel provided specific recommendations on how a contingent-valuation study should be designed and conducted to develop “reliable” estimates of passive-use values. The panel’s recommendations set off another wave of research designed to investigate the credibility of contingent valuation, particularly in the context of applying the NOAA Panel’s recommendations to the estimation use, passive-use values, and total values.[2]

From 1993 when the NOAA Panel report was published through 2003 when the first edition of this chapter was published, the contingent-valuation literature was populated with methodological studies designed to better understand the relative strengths and weaknesses of this valuation method in the context of the Panel’s recommendations. Moving forward from 2003 to the present, a review of the contingent-valuation literature still reveals methodological studies, but there are many more applications to a variety of issues around the world. Johnston et al. (2017) present contemporary guidance for stated-preference studies, including contingent valuation, that addresses all common applications, not just the estimation of passive use values for litigation.

Recently, the Journal of Economic Perspectives published a special section on contingent valuation as an update on what has been learned in the past 20 years. Kling et al. (2012) concluded that a “carefully constructed number(s) based on stated preference analysis is now likely to be more useful than no number,” and “the remarkable progress that stated preference researchers have made … serves as a model for the evaluation of other co-critical techniques” (p. 23). In contrast, Hausman (2012) continued to focus on the issues of hypothetical bias, disparity between willingness to pay and willingness to accept, and embedding/scope problems.[3] An element that is not clearly addressed in either of these papers is the standard that should be used to judge the credibility of contingent-valuation estimates. Kling et al. laid out four criteria where the strongest—criterion validity—asks if “the estimate generated by stated preference methods [is] the same as a willingness-to-pay value that would be generated if real payment was made” (p. 14). The question that is not well addressed is whether real payments are the

[1]The term “item” is used generically to refer to the object being valued. Due to the diversity of potential contingent-valuation applications, I did not want to implicitly restrict discussion to imply that studies are restricted to goods or bads, or commodities or resources, or quality or quantity, etc.

[2]The use of contingent valuation to estimate use values, which was the focus of the Bishop and Heberlein (1979) study, was not called into question. Some people have subsequently maintained that use-value studies must also meet the conditions set out by the panel, and some researchers have investigated the panel’s recommendations in the context of estimating use values.

[3]Hypothetical bias and scope are discussed in Sect. 4.2.2 in the discussion on validity. The discussion in this chapter focuses on estimating willingness to pay. Readers should see Chap. 2, Sect. 2.1.6 for a discussion of willing to pay and willingness to accept.

conceptually correct counterfactual to judge the credibility of contingent-valuation estimates and whether market contexts can affect real payments as survey design features might affect contingent-valuation estimates. Chapter 4 does not delve into these issues of validity, but some of these issues are covered in Chap. 12.

This chapter provides an overview of the steps in designing and implementing a contingent-valuation study. This chapter does cover the major considerations in designing and implementing a contingent-valuation study. Careful attention to the design features discussed in this chapter can enhance the overall credibility of any contingent-valuation study.

Valuation of groundwater quality is used as an example to help explain the steps involved in designing and conducting a contingent-valuation study. As contingent valuation is applied to a wide array of issues, the design features presented in this chapter are broadly appropriate for all applications. Methodological citations will be drawn from all types of applications published in the peer-reviewed literature, not just groundwater valuation applications. Design features unique to certain applications are not discussed here.

4.1 Steps in Conducting a Contingent-Valuation Study

Most of the action in designing a contingent-valuation study occurs in the development of the survey instrument in terms of scenario wording, choice of a valuation question, and bid selection where appropriate (Table 4.1). The valuation scenario in the survey is the foundation for value estimation, and the auxiliary questions in the survey provide respondent-specific data for statistical analyses. As choices in the survey design and data analysis stages of a study can affect welfare estimates, careful design of a contingent-valuation survey and careful analysis of the resultant data are crucial to the estimation of credible welfare estimates.

4.1.1 Identifying the Change in Quantity or Quality to Be Valued

The first step in conducting a contingent-valuation study involves developing a theoretical model of the value(s) to be estimated, which is based on the difference between the baseline utility with the current environmental condition and the utility with the new environmental condition (Step 1 in Table 4.1).[4] Following the notation protocol established in Chap. 2 (see Eq. 4.2), the value for a program to protect groundwater from contamination can be defined as

[4]"Change" is used as a generic reference to an increase or decrease in quality or quantity valued.

Table 4.1 Steps in conducting a contingent-valuation study

Step 1	Identify the change(s) in quantity or quality to be valued.
Step 2	Identify whose values are to be estimated
Step 3	Select a data collection mode
Step 4	Choose a sample size
Step 5	Design the information component of the survey instrument
Step 5.1	Describe the item to be valued
Step 5.2	Select a provision mechanism
Step 5.3	Select a payment vehicle
Step 5.4	Select a decision rule
Step 5.5	Select a time frame of payment
Step 5.6	Include substitutes and budget constraint reminders
Step 6	Design the contingent-valuation question
Step 6.1	Select a response format
Step 6.2	Allow for values of $0
Step 6.3	Address protest and other types of misleading responses
Step 7	Develop auxiliary questions for statistical analyses of valuation responses
Step 8	Pretest and implement the survey
Step 9	Analyze data
Step 10	Report study

$$v\left(P^0, Q^1, y\right) = v\left(P^0, Q^0, y - \mathrm{WTP}\right), \tag{4.1}$$

where $v(\cdot)$ is indirect utility function, P is the price of drinking water obtained from groundwater, y is income, WTP is a Hicksian measure of value, and Q is groundwater quality (Q^0 denotes current water quality and Q^1 denotes degraded quality if the protection program is not implemented).[5,6,7] The value of interest is that of protecting groundwater from contamination ($Q^0 > Q^1$). Note that P is assumed to remain constant. Assuming the change does not affect drinking water (e.g., water utility purification protects drinking water supplies), the value that would be estimated here is a passive-use value associated with a change in groundwater quality (ΔQ). The purpose here is not to delve into details of theoretical definitions of changes in welfare, which is the domain of Chap. 2, but to

[5]In Chap. 2, Flores uses C to designate WTP, and here WTP is directly used as the symbol.

[6]All other arguments in the indirect utility function are assumed to be constant between the status quo (Q^0) and the condition with diminished groundwater quality (Q^1), orthogonal to p and Q, and are suppressed for notational convenience. It is assumed here that some activity is undertaken to remediate the contamination such that the price of providing potable water increases ($p^0 < p^1$), but the activity is not fully effective so there is a residual value to be estimated.

[7]"Willingness to pay" is used in the presentation of this chapter while fully recognizing that many decisions actually require estimates of willingness to accept (Chap. 2, Eq. (2.3)).

discuss aspects of the value definition that must be considered in the design of any contingent-valuation study.

The theoretical definition of value is fundamental to three key components of any contingent-valuation study. First, the definition of value plays a central role in developing the description of the conditions with and without the item to be valued (Step 5.1 in Table 4.1).[8] Second, the definition frames the statistical analysis of the contingent-valuation responses (Step 9). Third, having a theoretical definition allows for clear interpretations of value estimates in specific decision-making contexts (Step 10). Decision-making is used in a very general sense to refer to any condition that can change utility and for which welfare estimates are required. This broad definition includes any public or private decision-making process where nonmarket values would be used.

The difficult part of Step 1 is identifying the physical change and how it affects utility. Economists depend on physical and biological information to establish status quo conditions (Q^1) and predictions of future conditions to be avoided (Q^0) or in other cases attained. This information may not be known with certainty, and applied valuation further requires information about this uncertainty. For example, a definition of value for a policy that reduces the probability that groundwater will become contaminated is as follows:

$$\begin{aligned} &\pi_1 v\left(P^0, Q^1, y - op\right) + (1 - \pi_1) v\left(P^0, Q^0, y - op\right) \\ &\quad = \pi_0 v\left(P^0, Q^1, y\right) + (1 - \pi_0) v\left(P^0, Q^0, y\right), \end{aligned} \tag{4.2}$$

where π is the probability of contamination, $\pi_0 > \pi_1$, and op is willingness to pay (option price as defined in Eq. 2.32, Chap. 2) to reduce the probability of contamination (Bishop 1982). Here, π_0 is the probability that the current water quality, Q^1, will not be degraded in the absence of a policy, and π_1 is the enhanced probability that Q^1 will not be degraded if the policy is enacted. Thus, op is maximum WTP to increase the probability that the current water quality will be maintained. Applying contingent valuation to conditions that change the probability of an outcome is a common practice.

The fundamental information transmitted in the survey is the description of the change to be valued. For example, if a policy will protect groundwater from contamination, the survey will explain current conditions in the aquifer (Q^1) and what would happen to quality if the protection policy were not implemented (Q^0).

Describing changes in quality in contingent-valuation surveys is an area where there has been considerable progress in recent years in two dimensions. First, there have been advances in the information on environmental quality available to study designers. These advances have evolved from interdisciplinary research projects where economists are interacting with physical and biological scientists at very early stages in the research process (Johnston et al. 2012), the expanding availability of GIS data (geographical information systems; Bateman et al. 2002), and the use of simulated images to portray new conditions (Madureira et al. 2011). The

[8]"Item" is used as a generic term to reference a quantity or quality change valued.

second evolution is the use of Internet surveys where the presentation of information can be carefully presented to respondents, such as allowing respondents to reference information as they complete the survey and limiting respondent's ability to read ahead or change answers as they complete the survey (Lindhjem and Navrud 2011). These advances have improved the realism of the information in contingent-valuation scenarios to respondents and helps to more carefully link value estimates to decision-making.

Sometimes, in the absence of clear information on the change to be valued, survey designers will focus on the action that affects the change in services people enjoy. With only a description of the action and not a description of the change, survey respondents are left to make two assumptions: what will be the change to value and how this change will affect the services they receive. This issue has not been directly and extensively addressed in the contingent-valuation literature and deserves more consideration. The key design consideration is to avoid descriptions that are vague or confusing. In the absence of clearly understood descriptions of the change to be valued, survey respondents are left to their own assumptions regarding what the change will accomplish. This is a problem because different respondents could make different assumptions about what to value, and those assumptions might not be consistent with the change the study is designed to value.

It could be the case that a contingent-valuation study is being conducted in advance of an anticipated decision, and the details of the action and the consequent effects are not known. Contingent-valuation studies are often designed to estimate an array of values that represent plausible conditions that might occur when the physical information becomes known or the actual decision is finalized. Thus, the design of a contingent-valuation study proceeds using a "best knowledge" of the actual change that will be implemented and of the ultimate effects with and without the action. Here is where a careful theoretical definition of value estimated is crucial; the definition allows an ex post interpretation of whether the value estimate matches the realized change it is intended to value.

In addressing this step, it is important to recognize that there are two types of contingent-valuation studies. Many studies in the peer-reviewed literature focus on methodological contributions and are not designed to address a specific decision-making issue (e.g., Murphy et al. 2010), while other studies have been designed to address a specific decision-making issue (Whittington and Pagiola 2012). This difference is crucial because methodological studies can use whatever definition of value is convenient. In a specific decision-making context, however, the value definition must be linked to the specific change in utility that will occur if the proposed change is implemented. Methodological studies still require theoretical definitions of estimated values to guide data analyses. Theoretical definitions of value can also allow estimates from methodological studies to be useful for benefit transfers (see Chap. 11).

The discussion in this section has focused on Hicksian measures of value, which is the theoretical concept typically measured in contingent-valuations studies. Step 1 involves an explanation of the theoretical definition of value presented within the survey instrument and the change that respondents are asked to value.

4.1.2 Identify Whose Values Are to Be Estimated

Once the change to be valued has been specified, the affected population can be identified (Step 2). This step involves identifying the affected population and selecting a sampling procedure (see Chap. 3, Fig. 3.1). This information is important in selecting a sample frame so that the contingent-valuation survey will be administered to a representative sample of affected individuals. It is also important to determine if the contingent-valuation study will estimate values on a per-capita or per-household basis. It is necessary to know the size of the affected population to expand the per-unit values to a population value. The affected population and unit of measurement could indicate a desirable mode of data collection or modes of data collection that are acceptable for the particular application.

In the example where groundwater quality is threatened, the affected population might constitutes those individuals who use the aquifer as a source of drinking water and those who hold passive-use values for groundwater in the aquifer. If everyone in a community obtains water for household usage that is drawn from the aquifer, it is fairly clear that the affected population includes everyone in the community. However, the affected population might also include people who live outside of the city and hold passive-use values. In this case it may be more difficult to identify the population holding passive-use values, and a uniform list might not be available to draw a sample from.

Most studies use geopolitical boundaries such as a city, county, or state in the United States to define the relevant study population. In the above example, the study population might be extended from the specific city to the county where the city is located if the aquifer of interest extends beyond the city boundary. The literature provides little guidance for selecting study populations (those who are affected by the change), but two points of consideration are useful. First, geopolitical boundaries are useful for identifying locations affected by the change being valued and those who will pay for the change to be implemented. Second, a spatially referenced sample will allow for an investigation of how value estimates change with distance from the affected area. If values decrease or increase with distance, this can affect the magnitude of aggregate welfare calculations (Bateman et al. 2006; Hanley et al. 2003; Mansfield et al. 2012).

Chapter 3 distinguishes between the affected population and the sample frame (those who are eligible for selection in the sample). Sample frame is important because this determines whether the selected sample is representative of the affected population.

Another issue is the unit of measurement for values. Some contingent-valuation surveys elicit values for individuals (e.g., Bateman et al. 1995), and others elicit household values (e.g., Poe and Bishop 2001). It is important that the framing of contingent-valuation questions makes clear whether a household valuation or an individual valuation is being requested. Mitchell and Carson (1989) asserted that "payments for most pure public goods are made at the household level. … When this is the case, the appropriate sampling procedure is to allow any adult who claims to be a household head to be a spokesperson for the household—the current practice of the U.S. Census Bureau" (pp. 265-266). (Mitchell and Carson cited Becker's (1981) *Treatise on the Family* in support of this assertion.)

However, the issue is more complicated. Quiggin (1998) argued that only when intra-household altruism does not exist or is paternalistic will the sum of individual values equal household values, yielding the same aggregate measure of welfare change. In the absence of these conditions, he argued that sampling and aggregating over individuals will yield larger measures of aggregate values than sampling and aggregating over households. In contrast, Munro (2005) argued that individual and household values will be the same only if household members pool their incomes. Supporting these potential limitations, Bateman and Munro (2009) and Lindhjem and Navrud (2012) found significant differences between individual and household values. These results would appear to violate the unitary model where households act like a person and there is only one utility function. However, Chiappori and Ekeland's (2009) collective model assumed income sharing among household members where it is possible to isolate individual welfare measures based on an allocation of resources to household members. In this theoretical framework, the potential contradictions mentioned above may not, in fact, be contradictions.

The lesson here is that restrictions on household member utility functions and interactions are necessary to secure equality between the sum of individual household member values and household values, and these assumptions might not always be testable. Further, it is not appropriate to assume that household values are a conservative approach. Thus, when choosing individual or household sampling units, it is important to ask several questions:

- Do households make individual or group decisions?
- If they make group decisions, do households pool their income?
- If they make group decisions, do decision-makers adequately account for the preferences of all household members?
- Do some households make group decisions and others make individual decisions?

Answers to these questions have important implications for individual and aggregate welfare estimation. A decision on individual versus household units of measurement should be based on the specific decision-making context being valued, and this selection influences what payment vehicle is selected (see Sect. 4.1.5.3), pretesting the survey instrument, the availability of a sampling frame, and other dimensions of the study design. A recent study suggests that people in a household act similarly, so this might not be an issue if the finding can be generalized to other applications (Adamowicz et al. 2014).

The bottom line for both sample frame and sampling units is that the sample frame should be known, and each unit in the sample should have a known probability of selection if value estimates are to be used to support decision-making.[9] Sample frame and unit of observation decisions should be clearly documented and any limitations or uncertainties noted in reporting study results.

4.1.3 Select a Data Collection Mode

A contingent-valuation study requires the collection of primary data (Step 3). The various modes of data collection are discussed in detail in Chap. 3 of this volume, so this section focuses on insights for contingent-valuation applications.

The most common way to implement contingent-valuation surveys is by mail (Schneemann 1997). Recently, Internet survey implementation has increased. Each method has its relative strengths and weaknesses (Dillman 2007). For a discussion of the alternative survey modes, see Chap. 3.

The primary reason that mail surveys are commonly used is that they have been the least expensive way to collect data. A recent call to a premier survey research firm suggests that the cost of a completed survey ranges from less than \$25 for Internet implementation to \$50–75 for a mail survey, \$75–150 for a telephone survey, and \$1,000–2,000 for personal interviews. These numbers are for national probability samples where an existing sample frame is assumed for the Internet implementation, and a national area probability sample is used for the personal interviews. Thus, with a limited budget, it is easy to see why Internet surveys are gaining popularity.

Another factor that has traditionally affected the choice of a survey mode is the expected survey response rate and the potential for item nonresponse to key questions in the survey (Messonnier et al. 2000). Even with careful design, Schneemann (1997) showed that response rates to contingent-valuation surveys are affected by factors that are not under the control of the study design. For example, decisions that involve specific user groups (e.g., anglers or hunters) are likely to result in higher response rates than general population surveys. Schneemann also found that survey response rates are related to the study application, the affected population, and the components of the contingent-valuation question itself, but this is not unique to contingent-valuation surveys. Rather, it is an issue to consider in the design of any survey (Groves and Peytcheva 2008). Thus, even with a good design process, it is important to recognize that features of the study application, as well as the design of the contingent-valuation component of the survey, affect survey response rates across all survey modes.

Response rates have been an issue of concern for Internet surveys. A credible Internet survey requires recruitment of individuals to represent the demographics of a known population, (e.g., the population of the country where the sample is drawn

[9]If the only purpose of a study is experimentation, the only important criteria may be random stratification, and sample representativeness generally is not a concern.

[see Chap. 3]). After this selection, the survey response rate can be less than 10% (Boyle et al. 2016; Manfreda et al. 2008). However, focusing on survey response rates as an indicator of survey quality and nonresponse bias has been shown to be misleading (Groves and Peytcheva 2008; Keeter et al. 2000). As a consequence, other measures of survey representativeness have been proposed (Groves et al. 2008; Schouten et al. 2009). Thus, while the expected survey response rate should be a design consideration, it is not necessarily a good metric to use to select the appropriate survey mode for implementing a contingent-valuation survey.

Conveying information on the item being valued is the fundamental component of a contingent-valuation survey. Personal interviews may be the most effective survey mode for conveying information because visual information can be provided and an interviewer is present to explain the information and answer questions.[10] An Internet survey allows for control of the information respondents read and can provide live links if a respondent wants to review information already presented, but an interviewer is not available to assist, answer questions, and monitor administration of the survey. A mail survey is more limited because no interviewer is present to explain the visual information and no control is provided on how respondents proceed through the survey instrument. The ability to provide information in a telephone survey is much more limited because no visual information is available. A mixed-mode survey where a telephone interview is conducted after respondents have received written and visual information in the mail, similar to the content of a mail survey, is one way to overcome the informational deficiencies of telephone interviews (Hanemann et al. 1991).

Other methods of implementing contingent-valuation surveys include on-site intercepts (Boyle et al. 1994) and convenience samples of students or other groups who are brought to a central location to participate in the study (Cummings and Taylor 1998). Davis and Holt (1993) noted that "behavior of (real) decision makers has typically not differed from that exhibited by more standard … student subject pools"[11] (p. 17; see also Smith et al. 1988). Maguire et al. (2003) found that college students behave similarly to a sample of all adults. These alternative survey modes with convenience samples are best used for implementing methodological experiments and not for studies designed to estimate values generalizable to a specific population to support decision-making. When experiments are conducted, it is still important that study participants are exogenously selected, and people are not allowed to voluntarily opt into the sample.

It is important to consider the advantages of each mode when implementing a contingent-valuation survey. A mail survey does not always dominate just because of cost advantages, and personal interviews do not always dominate just because of

[10]Personal interviews can also result in interviewer bias in responses to the contingent-valuation question. While Boyle and Bishop (1988) found interviewer effects with graduate students doing the interviewing, these effects might not persist if professional interviewers are used.

[11]Davis and Holt (1993) were referring to induced-value studies, and questions remain regarding whether contingent-valuation estimates derived from samples of convenience can be used to develop population inferences of gains and losses in economic welfare.

informational advantages. In reviewing studies that compared contingent-valuation studies implemented via the Internet versus mail, phone, and in-person implementation, Boyle et al. (2016) found that the ratio of willingness-to-pay estimates based on Internet surveys to estimates based on data from other modes averaged 0.92.[12]

4.1.4 Choose a Sample Size

Concerns about sample sizes (Step 4) are important for two reasons. First, the precision of estimated values affects their usefulness in decision-making. An estimate with large error bounds can leave questions about whether benefits really exceed costs or what a specific damage payment should be if a study were used to assess damages. Second, statistical precision affects the ability to detect differences among value estimates in methodological studies designed to investigate the validity and reliability of contingent-valuation estimates.

As noted in Chap. 3, selection of a sample size (Step 4 in Table 4.1) is a matter of choosing an acceptable level of precision within a given budget. That is, the standard error of mean WTP is

$$\mathrm{se}_{\overline{\mathrm{WTP}}} = \frac{\sigma}{\sqrt{n}}, \tag{4.3}$$

where σ is the standard deviation and n is the number of completed surveys used in the analysis. Thus, for a given variance, the absolute value of the standard error can be reduced by increasing the sample size. The larger is σ, the larger the required sample size to attain the desired level of precision. Additionally, for a fixed σ, a sample size will represent a larger percentage of the population for a small population than it will for a large population (Salant and Dillman 1994).

For applications where there have been a lot of contingent-valuation studies conducted (e.g., recreational fishing), there may be sufficient information to develop a reasonable estimate of σ for a new application. Where existing studies are not available, an estimate of σ can be obtained through a field test of the survey instrument. In practice, most studies choose the largest sample size possible given the available budget.

Selecting a sample size also involves consideration of the expected response rate and the expected item nonresponse to the contingent-valuation question and other

[12]Caution is warranted when interpreting these ratios as the comparisons are confounded by differences in sample frames and survey modes for all but one study. In the sole study where this confound was not present, the ratio was 0.86 (Linhjem and Navrud 2011); for the Boyle et al. (2016) study, which also controlled for sample frame, the ratio was 0.78.

variables that will be used to analyze contingent-valuation responses.[13] Other considerations include whether the sample will be stratified into subsamples for analysis and data reporting.

4.1.5 Design the Information Component of the Survey Instrument

Step 5 in Table 4.1 focuses on the information provided to respondents in the survey instrument. This includes telling respondents what it is they are being asked to value, how it will be provided, and how they will pay for it. Economics does not provide guidance for selecting these design features, and careful consideration must be given to these design features in the pretesting of the survey instrument. The design of the valuation scenario should be developed so that respondents believe the change to be valued can be feasibly accomplished and that the features selected do not unduly influence welfare estimates.

There are no hard and fast rules in the elements of this portion of the design phase. Thus, the use of cognitive interviews, focus groups, and pilot surveys is crucially important to understanding respondents' comprehension of the information presented and how they react to the information.

4.1.5.1 Describe the Item to Be Valued

The description of the item to be valued (Step 5.1 in Table 4.1) tells respondents what is being valued and presents the change in the quantity, quality, or probability to be valued (Step 1). The information set should include a description of the item being valued, baseline conditions, and the new conditions. This does not require an extensive, detailed valuation scenario, but a scenario that is clear so that respondents understand the change they are being asked to value.

The information is presented in written or verbal form and is accompanied by graphs, pictures, and other visual stimuli to facilitate respondent understanding. The information scenario is not a marketing or sales pitch but a neutral and fair description of the change to be valued, and it can include statements about why some people would desire the change and why others would not (Alberini et al. 2005). This information should also include an explanation of why the change is being valued.

[13]Consider a case where the desired sample size for statistical analyses is 500. Further consider that a mail survey will be used with a sample list that has 10% bad addresses, an expected response rate of 60%, and 10% item nonresponse to a valuation question. The initial sample size would need to be at least 1,029 [500 × (1/0.9) × (1/0.6) × (1/0.9)].

While the description of the item to be valued is the fundamental component in the design of any contingent-valuation study, it seems that the information is rarely complete in terms of the baseline condition(s) and the new condition(s) that will result as a consequence of the change (e.g., the levels of Q defined in Eqs. (4.1) and (4.2)). This problem arises most frequently in the estimation of values for marginal changes and to a lesser extent when total values are estimated. If respondents must infer the change being valued, it is likely that different respondents will use different subjective perceptions.

The appropriate quantity and quality of information required for respondents to provide valid value responses is a matter of much interest. While some have concluded that a credible contingent-valuation study requires that survey respondents be provided extensive, detailed information, the literature does not support such a conclusion. The literature does indicate that specific types of information should be provided to respondents. While only a small number of studies have investigated the effects of information on contingent-valuation estimates, these studies collectively tell an important story.

Samples et al. (1986) confirmed the obvious: You must tell people what it is they are being asked to value. Boyle (1989), in a study of trout fishing, found that providing respondents with additional information beyond the basic description of the change to be valued decreased estimates of central tendency, but the reductions were not statistically significant. The standard errors of the welfare estimates, however, decreased significantly with additional information. This result suggests information also affects the efficiency of value estimates. Bergstrom et al. (1990) investigated the effect of providing information about the services provided by wetlands and found that value estimates were affected by different types of service information. This outcome suggests that specific information on services is important for applications where respondents are not fully aware of how they currently benefit or could benefit in the future from the item being valued. Poe and Bishop (1999) demonstrated that specific information on well-water contamination was required in a study of groundwater protection. These findings are consistent with a more recent study by MacMillan et al. (2006) that found more information is required when people are less familiar with the item being valued. Valuation of a trout fishery by licensed anglers is an example where the study participants would be more familiar with the item being valued, but wetland and groundwater valuation are examples where study participants would likely have less knowledge or experience. These studies clearly indicate that specific information about the item being valued is required in order to elicit credible responses to contingent-valuation questions, and this is done through the qualitative research process to design the survey instrument.

Li et al. (2014) discussed prior knowledge and acquired knowledge people use in answering a survey. Prior knowledge includes knowledge people possess prior to engaging in the survey and any information they seek external to the survey while completing the survey (e.g., doing an Internet search for additional information). Acquired knowledge is the information provided in the survey instrument. Berrens et al. (2004) suggested that when given the opportunity to seek additional information, respondent use of this information is modest. This suggests that respondent attempts to seek additional prior knowledge could be limited. However, any additional prior information respondents seek introduces potential noise into welfare estimation. Thus, pretesting the information in the survey is needed to establish that survey respondents have the information they need to respond to the value question(s).

While some studies have used pictures and other types of graphics (e.g., maps, graphs, and tables) in valuation scenarios, there do not appear to be any studies that have evaluated whether the use of graphics to display information in the valuation scenario affects value estimates. The use of graphics requires careful pretesting so that the images do not inadvertently introduce unwanted effects into valuation responses. Thus, while pictures and other graphics can be helpful in conveying information in a contingent-valuation survey, they can also generate unwanted effects. The use of multiple modes of portraying information in a contingent-valuation scenario, such as written text, numerical presentation, graphs, or pictures, can facilitate respondent understanding because different people may use different information to understand the change to be valued.

Collectively, the lesson is that respondents to a contingent-valuation survey need to be presented with information that clearly explains the change to be valued and that such information must account for heterogeneity in how respondents use and process information. There is a careful balance between providing too little information such that respondents could misinterpret the valuation question and providing too much information so that respondents do not attend to critical portions of the information provided. Refinement of this information occurs in focus groups, one-on-one interviews, and, if necessary, in small-scale field pretests.

4.1.5.2 Select a Provision Mechanism

In any contingent-valuation study, it is necessary to tell respondents how the change to be valued will be implemented (Step 5.2). Without such information, respondents might not view the change as being credible or might implement a personal assumption that is not appropriate and could inadvertently affect value estimates. In a general sense, the provision mechanism is the "production process" that will accomplish the change respondents are asked to value. Suppose a policy was protection of well water from contamination. One provision mechanism that has been used to provide such protection is to establish protection zones around wellheads, which preclude any activities that might contaminate the groundwater. Sometimes applications have a clear mechanism that is part of the actual action being valued, while in other applications the selection of the provision mechanism is part of the study design.

Choosing the provision mechanism is complicated because the chosen mechanism could affect responses to contingent-valuation questions. For example, consider public concern over chemical residues in fruits and vegetables, genetically modified foods, sweatshop production of clothing, dolphin-free tuna, etc. These production attributes affect purchase decisions for market goods (Foster and Just 1989; Teisl et al. 2002) and there is no reason why there should not be similar types of provision-mechanism effects in responses to contingent-valuation questions. For example, while sweatshop production might not affect the quality of the clothes in terms of their use by the purchaser, this production process could represent an undesirable externality.

The effect of the selected provision mechanism on welfare estimates has not been formally investigated to my knowledge. At a minimum, careful pretesting in focus groups can identify whether respondents will understand the provision mechanism and if they feel it is credible in producing the change to be valued.

4.1.5.3 Select a Payment Vehicle

A payment vehicle (Step 5.3) is the mechanism by which respondents are told how payments would be made. For example, this might be a tax increase for a public policy or a higher price for a health-enhancing procedure.

This is a design area where the trade-off between credibility and unintended effects has been clearly noted in the literature. Mitchell and Carson (1989) argued that the choice of a payment vehicle requires balancing realism against payment vehicle rejection. That is, as realism increases, the likelihood that the payment vehicle will engender responses that protest the vehicle might also increase. For example, water-use fees are very realistic payment vehicles, but someone who values protecting potable groundwater might still give a valuation response of $0 to protest an increase in water rates. Income tax vehicles can run into problems due to resistance to higher taxes. On the other hand, a donation payment vehicle could yield an underestimate of value because this vehicle might not be incentive compatible for estimating a respondent's full willingness to pay (Wiser 2007).[14]

Failure to provide a realistic payment vehicle can also lead to protest responses. A sales tax would not be credible in an area that does not have sales taxes or when the group providing the item being valued does not have taxing authority. Thus, respondents could reject the valuation scenario even if they value the change because the payment mechanism is not believable. The realism of some vehicles can lead people to give what they think is a reasonable response, not their maximum WTP. For example, where there are only nominal entry fees (e.g., entrance fees at national parks), an increase in an entrance fee could engender responses of what respondents think is a reasonable increase in price rather than statements of maximum WTP (Campos et al. 2007).

Some studies demonstrate that payment vehicles do influence welfare estimates (Rowe et al. 1980; Greenley et al. 1981; Campos et al. 2007), but the line of inquiry

[14]Incentive compatibility refers to whether elements of the survey design will motivate people to truthfully reveal their preferences (i.e., statements related to the value they place on the item being valued).

Table 4.2 Payment vehicles used in recent studies

Payment vehicle	Study citation
Higher prices	Desaigues et al. (2011)
Voluntary donation	Garcia-Llorente et al. (2011)
Annual tax	Lindhjem and Navrud (2011)
Network fee	Menges and Beyer (2014)
Water bill	Ramajo-Hernández and del Saz-Salazar (2012)

has not been prominent in recent years.[15] Testing of payment vehicle effects is typically undertaken in survey pretesting to select a payment vehicle that minimizes undesirable effects on value estimates.

A variety of payment vehicles have been used in studies and a sampling of those in the recent literature are presented in Table 4.2. These examples are presented as a general set of payment vehicle examples and each could have relative strengths or weaknesses, as discussed above. While research can generate general insights about payment vehicles (e.g., donations are not incentive compatible and taxes are likely to lead to protest responses), selection of a specific payment vehicle is likely to be study-specific and will always require careful pretesting.

A concern with using prices or taxes as a payment vehicle is that people can adjust the quantity purchased (e.g., take fewer recreational trips or consume less water). This makes the cost to the individual endogenous; the respondent controls cost by adjusting quantity. Further, the choice of a payment vehicle must align with the value to be estimated. For example, using an increase in water rates to estimate passive-use values for groundwater (Eq. (4.1)) might not be logical to respondents and could result in some respondents including use values in their valuation responses. The concerns discussed here highlight how important it is to carefully select a payment vehicle that minimizes unintended effects on value estimates.

4.1.5.4 Select a Decision Rule

The decision rule is the process by which the results of the contingent-valuation study, individual valuation responses or summary statistics on valuation responses, are used to inform the decision as to whether the item valued will be provided (Step 5.4). Such a decision rule might be that the item will be provided if at least 50% of respondents answer "yes" to a dichotomous-choice question.

The choice of a decision rule is closely linked to the payment vehicle. A referendum is clearly applicable when the issue relates to the provision of a public good, such as groundwater protection, and the payment vehicle is an increase in taxes. However, a referendum would not be applicable when dealing with use

[15]In a recent conjoint study, Johnston et al. (1999) demonstrated that the choice of a payment mechanism, particularly one that guarantees funding for a change, can influence welfare estimates.

values, such as recreational fishing, and the payment vehicle is an increase in individual trip costs. In this second example, the decision rule mag be if aggregate benefits exceed project costs.

The more complex and perhaps more important cases are the public good cases like the groundwater valuation example. A fundamental goal in the choice of a decision rule is to make a selection that is plausible to respondents and will entice truthful valuation responses. Following Carson and Groves (2007), it is also important that the decision rule is consequential, which means that payment is mandatory if the program is imlemented and there is a non-zero probability that responses to the survey will influence provision of the item being valued.

4.1.5.5 Select a Time Frame of Payment

This step describes the number and frequency of payments respondents will make (Step 5.5). For example, a valuation scenario might be posed where a new filtration system would be installed to remove contaminants from a public water supply. Values could be elicited as a one-time payment now or as annual payments over the lifetime of the system, say 20 years.

While this is another area where there is scant research, Stevens et al. (1997) showed that repeated payments yield statistically different estimates of WTP when compared with a lump-sum payment. More recently, Soliño et al. (2009) found no difference between bimonthly and annual payments when a dichotomous-choice valuation question was employed. These limited results suggest that the choice of a payment time frame must proceed with caution.

There is often a disconnect between the time frame of payment in a contingent-valuation question and the time frame over which survey respondents will enjoy the benefits of the change. Typical changes result in benefits that accrue over a number of years, and the time frame of the payment(s) is often much shorter (e.g., a one-time payment). Thus, survey respondents are asked to undertake personal discounting to answer valuation questions. The research by Stevens et al. (1997) suggested that survey respondents might not do this well.

The time frame of payment varies substantially across studies, from one-time payments to annual payments into perpetuity. The time frame is crucially important because this influences how value estimates are aggregated to compute benefits or costs. This is another design feature that must be carefully addressed in survey pretesting.

4.1.5.6 Substitutes and Budget Constraint Reminders

Substitutes and a budget constraint are fundamental components of economic choices (Step 6). Both the availability of substitutes (Freeman 1993, Chapter 3; Hoehn and Loomis 1993; Flores, Chapter 2 in this book) and income (Flores and Carson 1997) affect the magnitude of welfare estimates.

Though encouraging respondents to consider substitutes and think about their budget constraints when answering contingent-valuation questions is intuitively straightforward, it is difficult to test the effectiveness of these reminders. What might be considered a substitute by one respondent might not be considered a substitute by another respondent. Split-sample studies, where one sample is reminded of substitutes and their budget constraints and another sample is not, reveal that information on substitutes, complements, and budget constraints affect estimates of central tendency and dispersion (Kotchen and Reiling 1999; Loomis et al. 1994; Whitehead and Bloomquist 1995). In a meta-analysis, Schläpfer (2006) found significant income effects for contingent-valuation estimates. Smith (2005) suggested that sensitivity to budget becomes more relevant the higher the cost to the respondent is as a proportion of income. Given the roles that substitutes, complements, and income play in theoretical definitions of economic values, the theoretical component of content validity suggests that respondents should be prompted to consider likely substitutes and complements, and they should be reminded that they could spend their money otherwise.

4.1.5.7 Summary

There is no cookie cutter or one-size-fits-all set of rules for framing contingent-valuation scenarios, but testing draft information scenarios in focus groups is critically important. Focus group testing provides the opportunity to learn if respondents are using the information, understand and believe the information, and are basing valuation responses on the actual change being valued.

Even seemingly innocuous statements and terms in contingent-valuation scenarios have the potential to affect valuation responses. For example, in a study of preserving agricultural lands we quickly found that open space has entirely different meanings to people involved in land-use policy as compared to the general public. Focus group participants told us that open space conveyed a sense of "outer space," "large foyers," etc.—not undeveloped land.

Pretesting in focus groups and/or one-on-one interviews is the best way to avoid pitfalls that can bias welfare estimates because of incorrect interpretation of information by respondents, the provision of unintended clues to respondents, and information rejection by respondents. This pretesting must be carefully conducted and is not a substitute for more research to understand the effects of each element of the information in a contingent-valuation scenario. This means that careful design must also be accompanied by conceptual and methodological research to refine what and how information should be presented in the survey to guide the design of future empirical studies.

In practice, it is important to recognize that the information set must vary from study to study to fit the issue-specific application and institutions. Further, any information set will not be understood and accepted by everyone in a sample; the design goal is to minimize misinterpretations and scenario rejection to the greatest extent possible. This means that some information in the scenario will be necessary to satisfy some respondents and other information will satisfy other respondents; the common goal is to elicit consistent value information across all respondents. Little is documented about study pretesting, but reporting of this information in journal articles and other publication will help the valuation community learn about design challenges and successes from prior study designs.

4.1.6 Design the Contingent-Valuation Question

After the information regarding how the change to be valued will be provided, the respondents are asked to reveal information about the value they place on the change described in the valuation scenario. This section provides guidelines and considerations for selecting and designing a contingent-valuation question (Step 6).

4.1.6.1 Select a Response Format

The response format refers to how the contingent-valuation question will be answered (Step 6.1). The three main formats ask respondents to directly provide their maximum willingness to pay (open-ended), choose an amount from a list of possible willingness-to-pay amounts (payment-card), or respond "yes" or "no" to a specified dollar amount (dichotomous-choice). The response format has implications for how the response data are analyzed and interpreted; it is the key characteristic that differentiates the various types of contingent-valuation questions. The information scenario components described above are generally portable from one question format to another with the exception of the decision rule (e.g., a majority vote would work with a dichotomous-choice question but not with an open-ended question).

Early contingent-valuation studies used either an open-ended question (Hammack and Brown 1974) or an iterative-bidding question (Randall et al. 1974). An open-ended question asks respondents how much they would pay for the specified change. An iterative-bidding question starts by asking respondents, "would you pay $SB" for a specified change (SB = starting bid). If respondents answer "yes," then the bid is increased in specified increments (I) until they say "no" and decreased until they answered "yes" if the initial response was "no" ($\$SB \pm \$I$). The magnitudes of starting bids, magnitudes of bid iterations, and number of iterations varied from study to study. While the open-ended format has persisted, the iterative-bidding format is no longer used because of an anchoring effect where the final bid at the end of the iterations was found to be significantly correlated with the starting bid (i.e., the higher the starting bid, the higher the final bid to which people would answer "yes" (Boyle et al. 1985; Thayer 1981).

If the passage of the proposal would cost you these amounts **every year** for the foreseeable future, what is the highest amount you would pay and still vote for the program? (CIRCLE THE HIGHEST AMOUNT THAT YOU WO ULD STILL VOTE FOR THE PROGRAM)

10¢	50¢	$1	$5	$10	$20
$30	$40	$50	$75	$100	$150
$200	MORE THAN $200				

Fig. 4.1 Example of an unanchored payment card (Welsh and Poe 1998, p. 183)

Open-ended questions are still used in some studies. The following is an example of an open-ended question used by Welsh and Poe (1998):

> If passage of the proposal would cost you some amount of money **every year** for the foreseeable future, what is the highest amount that you would pay annually and still vote for the program? (WRITE IN THE HIGHEST DOLLAR AMOUNT AT WHICH YOU WOULD STILL VOTE FOR THE PROGRAM) (p. 183)

Respondents are provided with a blank line where they can write in the maximum they would pay.[16]

In the early 1980s, Mitchell and Carson (1981) introduced the payment card (see also Mitchell and Carson 1993). This was a card with k bid amounts, and it showed respondents how much they pay for selected public services (anchors), which in essence is very general information on substitutes. Respondents were asked to "circle the dollar amount that is the most they would pay" for the change. Current applications of payment cards have proceeded without anchors (Fig. 4.1).

Dichotomous-choice questions, introduced by Bishop and Heberlein (1979), ask respondents, "would you pay $B" for the specified change, which is simply the first round in an iterative-bidding question (Fig. 4.2). The bid amount ($B) is varied over different respondents. The starting-point problems with iterative-bidding questions subsequently led to the adoption of dichotomous-response questions, and the single-shot question is easier to administer than the iterative framework. Some have also posited the heuristic argument that dichotomous-choice questions mimic the take-it-or-leave-it nature of many market purchases. Such a heuristic argument cannot be made for open-ended and payment-card questions.

[16]A well-defined contingent-valuation question would also tell the respondents the period over which payments would occur.

Would you vote for this proposal if the proposal would cost you $B **every year** for the foreseeable future? (CIRCLE ONE NUMBER)
1 Yes
2 No

Fig. 4.2 Example of a dichotomous-choice question (Welsh and Poe 1998, p. 183)

This example is a dichotomous-choice response format framed as a referendum. The bid amounts ($B) are entered before the survey is administered.[17] The referendum vote here is the decision rule. A dichotomous-choice question can be framed as a referendum or not. For example, a dichotomous-choice question can be framed as agreeing to pay or not pay an entrance fee to a national park.

A number of researchers have experimented with variations of the dichotomous-choice format. For example, studies have used one-and-one-half bounded questions (Cooper et al. 2002), double-bounded questions (Hanemann et al. 1991), and multiple-bounded questions (Bateman et al. 2001). Each of these questions present follow-up bids to survey respondents. For example, in the one-and-one-half bound, respondents randomly receive an initial bid. If they answer "yes" to the initial bid amount, they receive a higher bid; if they answer "no," they receive a lower bid amount. The multiple-bounded question is a repeated dichotomous choice where a response is required for every bid amount, which is essentially a payment card where respondents indicate their willingness to pay each bid amount, not just the maximum they would pay. These alternative specifications of dichotomous-choice questions were proposed to increase estimation efficiency (Hanemann et al. 1991). Responses to a dichotomous-choice question only reveal if each respondent's value is less than ("no" response) or greater than ("yes" response) the bid amount they received. Adding additional bid amounts reduces the range into which the unobserved values reside.

One of the NOAA Panel recommendations (Arrow et al. 1993) was to allow respondents a "no answer" option in addition to "yes" or "no" when the valuation question was framed as a referendum. This recommendation appeared to logically follow from consideration of undecided voters in predicting election outcomes. While there have been many different interpretations of how "don't know" responses should be elicited and treated in data analyses, at the basic level recommended by the NOAA Panel, it appears that most these respondents would vote "no" in the absence of a "don't know" option (Carson et al. 1998; Groothuis and Whitehead 2002).

[17]Welsh and Poe (1998) used nine bid amounts that ranged from $1 to $200.

Payment-card questions and dichotomous-choice questions require an additional design feature of selecting the bid amounts used as the monetary stimuli in the questions. Development of bid values usually follows a three-step process. The first step is to review similar studies in the literature to develop a prior on the distribution of values for the current application. Second, this prior information is used to develop the initial bid amounts used in pretesting the survey instrument, and these bid amounts are adjusted based on what is learned in this survey design process. This should include a field pretest, or pilot, of the full survey instrument and should not be limited to focus group data if possible. Finally, an optimal bid-design approach can be used to select the bids amounts used in the final survey instrument (Alberini 1995a; Dalmau-Matarrodona 2001; Kanninen 1993a, b; Scarpa and Bateman 2000). Alberini (1995a, b), and Kanninen (1993a, b, 1995) have shown that an optimal design has a small number of bids (five to eight), and the bid amounts should span the median WTP, and not placed too close to the median nor in the tails of the distribution. Very low or very high bid amounts may not be credible to respondents. Mis-specification of the bid distribution such that the most bid amounts fall above or below the median seriously compromises the ability to estimate mean WTP. McFadden (1994) proposed a continuous bid design, which might avoid mis-specification when only a small number of bids are employed. These bid-specification issues were empirically investigated by Boyle et al. (1998), with the empirical results supporting the bid-design recommendations of Kanninen (1993a, b), and Alberini (1995a).

The framing of the actual contingent-valuation questions and their respective response formats are quite simple relative to the framing of the valuation scenario that precedes the valuation question. The one exception is that careful selection of bid amounts is crucial in the design of payment-card and dichotomous-choice questions. In the next section, the relative strengths and weaknesses of these question formats are discussed.

4.1.6.2 Relative Strengths and Weaknesses of Response Formats

While dichotomous-choice questions are most commonly used, each of the three main response formats has strengths and weaknesses (Table 4.3). Conceptual arguments by Carson and Groves (2007), Carson et al. (2014), and Hoehn and Randall (1987) suggest that the “take-it-or-leave-it” nature of dichotomous-choice

Table 4.3 Comparison of contingent-valuation response formats

Characteristics	Open-ended	Payment card	Dichotomous choice
Incentive compatible	No	No	Yes
Bid design required	No	Yes	Yes
Responses/statistical efficiency	Continuous	Interval	Greater than or less than a threshold bid
Potential problems	Zero bids	Anchoring	Anchoring

questions, when framed as a referendum vote for a public good, has desirable properties for incentive-compatible revelation of preferences. There is a single bid amount to which respondents respond, and there is no incentive for respondents to pick very high (more than they would pay) or very low (less than they would pay) dollar amounts to purposely misstate their values. This is not the case for open-ended and payment-card questions where respondents can influence the outcome of a study by the value they state or dollar amount they pick. For example, if respondents want to see a change occur, they can state an open-ended value or pick a payment card amount that exceeds their WTP. Alternatively, if they want to send a signal that they want the cost to be low, they might select a value below what they would actually pay. The opportunities for such misstatements of value are not consistent with incentive comparability.

Cummings and Taylor (1998) argued that dichotomous-choice questions must be accompanied by the realism that the referendum vote will be binding (i.e., respondents must believe the change **will** be implemented if more than 50% of respondents vote "yes"). This concept has been more formally developed by Carson and Groves (2007) and Vossler et al. (2012). In contrast to Cummings and Taylor (1998), Carson et al. (2014) argued that it is not necessary that the referendum is binding but that the results of the survey referendum will have a nonzero probability of being used in the decision-making process to provide the item being valued. This provides a strong incentive for dichotomous-choice questions framed as a referendum as the preferred framing of a contingent-valuation question.

Responses to open-ended questions result in a continuous distribution of responses on the interval $[0, +\infty)$, while payment-card responses reveal whether the respondents' values reside within a $k + 1$ interval where k is the number of bid amounts ($B) on the payment card $[\$B_L \leq \text{WTP} \leq \$B_U)$, where WTP is willingness to pay, B_L is the bid chosen by the respondent, and B_U is the next bid higher than the chosen bid. Responses to dichotomous-choice questions indicate only whether each respondent's values lie below $[\$0, \$B)$ or above $[\$B, +\infty)$ the bid threshold. Assuming truthful revelation of responses, a person with a value of $15 would respond in the following manner to each of the three basic contingent-valuation response formats:

- The response to an open-ended question would be "$15."
- The response to a payment-card question with bids of $1, $10, $20, and $30 would be "$10."
- The response to a dichotomous-choice question with a bid amount of $10 would be "yes."

Thus, the dichotomous-choice response format reveals if a respondent's value lies in the interval $[\$10, +\infty)$; for the payment-card response format, the respondent's value resides in a narrower interval $[\$10, \$20)$; and for an open-ended response format, the value of $15 is observed. Therefore, in terms of estimating central tendency, the open-ended format provides the most efficient estimates, while the dichotomous-choice format provides the least efficient estimates ($se_{oe} < se_{pc} < se_{dc}$, where se is the standard error of the estimated mean).[18] This relationship assumes that all three response formats incentivize respondents to truthfully reveal their preferences.

However, each of the response formats can have unique impacts on respondents' answers to a contingent-valuation question. For example, open-ended questions are believed to yield an unusually high percentage of responses of $0 in that some people might hold a value, but answer $0. It is also argued that people have difficulty coming up with a maximum willingness to pay amount for policies they are not familiar with. A manifestation of this issue is that the empirical distributions of responses to open-ended questions are not smooth and tend to have spikes at $5 increments. This rounding to the nearest $5 further attests to the difficulty respondents might have giving a precise dollar value. For examples of issues with open-ended questions, see Bohara et al. (1998), Donaldson et al. (1997) and Welsh and Poe (1998). All in all, very few applications use open-ended questions today.

Payment cards appear to avoid the issues of a spike of zero values and respondents having to provide a specific dollar value. However, Rowe et al. (1996) found that the bid amounts on the payment card can influence value responses. With careful framing, a payment card question can be posed in the context of a referendum and presented as consequential. Only a few studies still use payment cards (Covey et al. 2007; Ryan and Watson 2009), but despite this low usage, payment-card questions might be the best alternative to dichotomous-choice questions.

While dichotomous-choice questions gained popularity to avoid the anchoring in iterative-bidding questions, dichotomous-choice questions are not free from anchoring problems (Boyle et al. 1997, 1998; Green et al. 1998). That is, respondents have a propensity to say they would pay bid amounts that likely exceed their true values and to say they would not pay low bid amounts below their true values. The issue seems to be most problematic with high bids, which would serve to inflate value estimates (Boyle et al. 1998). Prices and quality are often perceived as being correlated in market goods, and this market intuition could lead respondents to interpret single bids as implicit signals of quality that lead to anchoring (Gabor and Granger 1966; Shapiro 1968).

Further, concerns about the effects of bid amounts on responses extend to one-and-one-half-bound, double-bound, and multiple-bound questions (Bateman et al. 2001, 2009; Herriges and Shogren 1996; Roach et al. 2002; Watson and Ryan 2007). Thus, while dichotomous-choice questions are theoretically incentive compatible, research suggests that value estimates might not be robust to manipulations in the bid design. An interesting question is whether a highly consequential survey would be less susceptible to bid effects.

[18]These denote the standard error (se) of the mean value estimated using the open-ended, payment-card and dichotomous-choice response formats.

Dichotomous-choice questions, posed as a referendum vote, are the safe approach to frame contingent-valuation questions, and the extensive use of this approach in the peer-reviewed literature supports this endorsement. However, the referendum framing of a dichotomous-choice question is not practical in some contexts such as recreation use values. Following are some additional considerations in the framing of contingent-valuation questions.

4.1.6.3 Allowing for Values of $0

Some people include the issue of zero bidders under the general heading of protest responses, but there are two issues here (Step 6.2). The first relates to people who give a response of $0 because they reject some component of the contingent-valuation scenario; these are protest responses that will be dealt with in the next section. This section considers those who truly hold values of $0. It is quite possible that a change might not be utility-increasing for some segment of the sampled population, and respondents need a way to indicate such a lack of value.

With an open-ended question, a respondent can simply enter a response of $0, and a payment card can include a value of $0 for respondents to circle. The more problematic case is a dichotomous-choice question where respondents can answer "no" to the bid but do not have the opportunity to express a value of $0. In these cases, we know only whether respondents' values lie within the interval (−∞, $B). We do not know if there is a spike in the probability distribution at $0, and it is necessary to have a separate question to identify respondents whose values are $0.[19] This $0-value screen question has been implemented by posing it before the contingent-valuation question and then administering the valuation question only to those who answer "yes" to this screen. Alternatively, this question could probe respondents after they have answered the contingent-valuation question by asking respondents who answer "no" to the bid if they would "pay anything" for the change.

For example, Ahearn et al. (2003) used the following question that preceded the contingent-valuation question: "Would you vote for the proposal if passage of the proposal would increase your household's 1998 income tax?" Respondents who answered "no" were not asked the contingent-valuation question.

A related issue is that policies might actually give some people disutility, which would imply that their values would be strictly negative. While it appears that most studies treat people with negative values as $0s or that such outcomes are artifacts of the statistical distributions assumed in econometric estimation (Haab and McConnell 1997), others have attempted to investigate the plausibility of negative values (Berrens et al. 1998; Bohara et al. 2001).

[19]A similar question is required for a payment card if a bid amount of $0 is not included.

4.1.6.4 Protests and Other Types of Misleading Responses

There are at least three types of potential response categories under the heading of protests, all based on a presumption that these are respondents who do not report their true values (Step 6.3). It is important to note that these can be overtly misleading responses or misleading responses that occur inadvertently. Inadvertent misstatements can occur because someone does not fully understand the valuation scenario or because of experimentally induced errors.

The first category includes people who protest some component of the contingent-valuation scenario. These respondents might answer "$0" even though they hold a value for the item, which biases the estimate of central tendency downward, or they might choose not to complete the survey, leaving the effect on central tendency dependent on how these respondents are treated in the analysis of the contingent-valuation data.

The second category includes people who do not understand what they are being asked to value and who answer the valuation question anyway. The effect of this misunderstanding might not introduce a bias into estimates of central tendency, but it most likely will increase noise in the data that will increase the standard error of the mean.

The third category is people who behave strategically in an attempt to influence survey results and ultimately the decision. If everyone who is behaving strategically acts in a similar manner, the effect will be to introduce a bias into the estimate of central tendency. However, some people could have incentives to understate values, and others could have incentives to overstate the values, leaving the overall effect on estimates of central tendency indeterminate.

Within the contingent-valuation literature, two types of misleading responses have received specific attention: warm glow and social desirability. "Warm glow" arises from the utility that people receive from stating a willingness to pay and not for the change actually being valued (Andreoni 1989). Some have suggested that warm glow confounds estimation of WTP, but perhaps the effects can be removed from estimates (Nunes and Schokkaert 2003). However, the warm glow literature has largely been developed for philanthropy and donations (Harbaugh 1998), and the extension to contingent-valuation estimation of WTP is not fully explored. Donations are not a desirable payment vehicle for contingent-valuation questions because the goal is to estimate willingness to pay at the point of indifference, which is shown by the equality in Eqs. (4.1) and (4.2). This is what some people refer to as "maximum willingness to pay." It is not a donation toward provision of the item being valued, but the measurement of a specific economic concept.

Social desirability bias arises when respondents answer questions in a manner to please another person such as the interviewer in a personal interview. While some have detected social desirability (Leggett et al. 2003) in contingent-valuation estimates, it is likely that this effect is limited to interview formats where there is a clear incentive to please, which is not the general case for contingent-valuation studies. For example, the Leggett study was conducted in person on a boat as people were returning from visiting a military historical site in an area people visit because of the military history. Even in cases where social desirability might arise, it can be addressed through survey implementation procedures (Kreuter et al. 2008).

In terms of identifying misleading responses, empirical applications have used a variety of approaches to identify anomalies in responses to contingent-valuation questions. Some have included questions in the survey to probe respondents' understanding and motivations when answering the contingent-valuation question (Ajzen et al. 1996; Berrens, Bohara et al. 1998; Blamey et al. 1999; Stevens et al. 1994). Others have trimmed the upper values if they are greater than a certain percentage (e.g., 10%) of a respondent's income (Mitchell and Carson 1989, pp. 226-227). Others have used statistical routines as described in Belsey et al. (1980) to identify responses that have undue influence on estimation results (Desvousges et al. 1987).

While most acknowledge that there are potentially some misleading responses in contingent-valuation data, there is no established procedure with a sound conceptual basis for excluding responses. The reasons for this are varied. What if a person gives one response that suggests a protest, but provides another response that indicates they are providing a valid response? Which response is correct or more meaningful? Or what constitutes a sufficient lack of understanding such that a respondent's valuation response should be excluded from statistical analyses? For example, it would not be appropriate to exclude respondents with low levels of education because they still have preferences and hold values. They might not be able to understand the valuation scenario as well as other respondents, but they make market decisions with a similar ability on a daily basis. Questioning people after they answered the valuation question is problematic; people who are behaving strategically would be unlikely to tell you that they are doing this. People who do not understand the valuation question also might not understand the follow-up question. In addition, these responses to follow-up questions cannot be assumed to be exogenous to responses to the valuation question.

Another approach, trimming the tails of the distribution by deleting outliers, must be done with care (e.g., people with high values might be those who have the most to lose). For example, some people give up income to live in areas that are near desirable resources; for these people, welfare losses could be quite large relative to income.

Another question deals with how much of an effect misleading responses actually have on estimates of central tendency. The famous Marwell and Ames (1981) study found that "economists free ride," but others do not. This suggests that strategic behavior might be relegated to a small segment of any sample. Thus, those who behave strategically or who more generally provide protest responses might be a small segment of the sampled population and might not behave in a way that influences sample statistics. The first contingent-valuation study that I conducted was undertaken with personal interviews of people while they were recreating on-site. An environmental group came through one of the survey locations; they talked among each other and encouraged each other to behave strategically by giving high-value responses to influence a desirable environmental outcome. We marked all of the surveys to identify these individuals in data analyses. Despite behaving strategically, none of their responses were statistical outliers and most responses were quite close to the sample mean; their strategic behavior was not effective at manipulating sample statistics. Thus, while strategic behavior could occur, it is possible that it is not sufficiently pervasive or of a magnitude to have an effect on welfare estimates.

Recent studies have looked at segmenting true and protest zero responses. Jones et al. (2008), based on an open-ended question, segmented true and protest zero responses and then analyzed the data using a Tobit model after removing protest zero responses. Strazzera et al. (2003) provided a more sophisticated econometric approach that accounts for both true and protest zero responses without any ex ante segmenting. Meyerhoff and Liebe (2006) looked at protest more generally throughout their data and found that a variable that scales the level of protest response resulted in a higher level of protest, decreasing the likelihood that someone would hold a nonzero value and decreased estimated willingness to pay.

In contrast to treating protest responses as faulty data, Garcia-Llorente et al. (2011) suggested that protest responses can be useful in identifying specific design elements of environmental programs that will engender greater public support. This changes the perspective and dimensionality of a contingent-valuation study from simply estimating a value to support decision-making to providing richer information to decision-makers regarding program design.

Despite the issues discussed above, contingent-valuation study designs should consider including questions to differentiate true $0 responses from potential protest $0s, investigate the presence of data outliers, and include questions to probe respondent acceptance of the change valued, the provision mechanism, the payment vehicle, the decision rule, the time frame of payment, and the actual payment amount in payment-card and dichotomous-choice questions. While the issue of misleading responses to contingent-valuation questions deserves consideration, it is a tough conceptual and empirical issue that deserves greater consideration at a conceptual level of what actually constitutes a protest response such that the observation should not be included in data analyses to compute value estimates. Further, additional robustness studies are needed to determine if there is a systematic effect on estimates of central tendency and dispersion.

4.1.7 Develop Auxiliary Questions for Statistical Analyses of Contingent-Valuation Responses

Step 7 calls for development of auxiliary questions designed to collect data to be used in the analyses of responses to the contingent-valuation questions, and auxiliary questions can provide information to support decision-making beyond what is provided by the valuation component of the study. These are in addition to probing for potential protest responses, which was discussed in the preceding section. The most obvious candidates are income (Schläpfer 2006) and other variables that might influence value estimates (e.g., Dupont 2004; Farreras et al. 2005).

Questions such as income, sex, and other demographic questions that have objective responses are commonly placed at the end of a questionnaire following standard survey practice. Opinion questions, whose responses are to be used in analyzing responses to the valuation question, might include a question asking respondents to rate the current condition of the item being valued. Such a question should be placed before the valuation question(s) in a survey. If such questions are placed after respondents have answered the valuation question, their responses cannot be assumed to be exogenous to responses to the valuation question. Data on these types of variables are commonly used in the estimation of econometric models of payment-card and dichotomous-choice data.

Secondary data may also be incorporated in econometric analyses. For example, in a study of lost value from contamination of an aquifer, Poe and Bishop (1999) argued that well-specific water quality data are needed in the valuation question. Spatial data can also be useful to develop variables that describe proximity of respondents' households to wells with known levels of contamination or proximity to the source of the contamination. It is important to have questions in the survey that will help in matching the respondent's answers to valuation questions to the auxiliary data that will be merged with the survey response data.

Finally, it is important to consider whether existing surveys (e.g., the U.S. Census, NORC General Social Survey, etc.) have similar questions. Using the same framing of questions as existing surveys allows for direct comparisons of the data for assessing whether sample selection has occurred, merging data sets for richer statistical analyses, and filling in missing data due to item nonresponse. Cameron et al. (1999) used such an approach to address nonresponse bias in a mail survey.

4.1.8 Pretest and Implement the Survey

Chapter 3 provides details of survey design and administration (Step 8). This section will briefly discuss the use of cognitive interviews, focus groups, and pilot surveys in the design of a contingent-valuation survey.

Surveys are pretested through an iterative process of one-on-one interviews, focus groups, and/or field trials (pilot surveys). Here again, there is no one-size-fits-all approach. Some studies might use one focus group, while other studies use multiple focus groups. Some studies might use focus groups and cognitive interviews, while other studies only use focus groups. These choices can depend on the available budget and the complexity of the valuation topic. For studies with a large budget and a complex valuation topic, the design process could be the following:

- Focus groups to learn about potential respondent's knowledge and beliefs about the item being valued.
- Focus groups to test the valuation scenario.
- Cognitive interviews to learn what people think about the valuation scenario in the absence of group effects and to learn more deeply about how potential respondents are reacting to the survey information.
- Focus groups to pretest a complete draft of the survey instrument.
- A field pilot to develop information on the survey response rate and item responses to individual questions in the survey with particular interest in responses to the valuation question.

How qualitative research tools and field pilots are used in the survey design process varies from study to study, but there is a common feature. The design process typically starts with a simple structure to learn from potential respondents, and the complexity of the survey instrument is built through what is learned in each focus group, cognitive interview, and pilot. While this qualitative research is common to survey design, the unique component here is the design of the contingent-valuation scenario.

The give and take of the focus group process can be useful by learning what is important to one person in terms of information about the item being valued is not very important to the rest of the group. Learning how different people in the focus groups respond to different features of the valuation scenario is a crucial feature of this survey design research. Cognitive interviews allow for more in-depth probing of potential respondents' reactions to the surveys materials in the absence of the group dynamics. The pilot survey is particularly useful for developing information on potential values to be used to develop bids for payment-card and dichotomous-choice questions.

This qualitative pretesting of the survey instrument and administration process ensures that survey questions are understandable to respondents and are actually eliciting the information they are designed to elicit, and that the survey will yield adequate responses to support statistical analyses of the resulting data.

Following the pretesting, the survey should be implemented using best practices for survey administration as discussed in Chap. 3.

4.1.9 Data Analysis

The process of data analysis (Step 9) varies with the response format used for the contingent-valuation question. We start with responses to open-ended questions, move to payment-card responses, and close with dichotomous-choice responses, thereby covering the three main types of contingent-valuation response formats.

Responses to open-ended questions are mathematically the easiest to analyze in terms of computing the arithmetic mean:

$$\overline{\mathrm{WTP}} = \sum_{i=1}^{N} \frac{\mathrm{WTP}_i}{N}, \tag{4.4}$$

where WTP_i is the open-ended response for the ith respondent, and N is the number of observations (complete surveys). The responses to an open-ended question (WTP_i) are individual statements of value, WTP or op in Eqs. (4.1) and (4.2), respectively. If open-ended responses are analyzed as a function of variables that explain WTP, a theoretical specification would be based on a definition of the value as in Eqs. (4.1) and (4.2), which would be solved for WTP or op as the dependent variable.

Analyses of payment-card and dichotomous-choice data require econometric analyses and the estimated equations are used to derive estimates of WTP. These analyses start with the specification of a function based on theoretical definitions of value, like the examples presented in Eqs. (4.1) and (4.2). Willingness to pay can be expressed as

$$\log(\mathrm{WTP}_i) = x_i'\alpha + e_i, \tag{4.5}$$

where $x_i^{'}$ is a vector of arguments that affect the magnitude of individual's WTP, α is a vector of preference coefficients to be estimates, and e_i is the random error that might be assumed to be distributed normally with mean zero and standard deviation σ. The function $x'\alpha$ is specified as the solution to equations such as (4.1) or (4.2) that define the value being estimated. The vector x_i logically includes variables that describe the change valued. As with most econometric analyses, the explanatory variables are often chosen based on economic theory and previous research, and are hypothesized to affect the magnitudes of respondents' WTP.

Analysis of payment-card data proceeds by modeling the interval where respondents have revealed that their values reside. These intervals are bounded by the bid amounts each respondent circled and the next highest amount on the payment card. Following Cameron and Huppert (1989), respondents' true values $\left(\mathrm{WTP}_i^t\right)$ reside in the interval ($\$\mathrm{B}_{1i}$, $\$\mathrm{B}_{ui}$], where "1" denotes the lower bid the respondent circled and "u" denotes the next highest bid amount on the payment card.[20]coefficients. Since WTP_i is not actually observed, the probability that WTP_i falls into the chosen interval on the payment card is modeled as

$$\Pr(\mathrm{WTP}_i \in (\$\mathrm{B}_{1i}, \$\mathrm{B}_{ui}]) = \Pr\left(\frac{\log \$\mathrm{B}_{1i} - x_i'\alpha}{\sigma} < t_i < \frac{\log \$\mathrm{B}_{ui} - x_i'\alpha}{\sigma}\right), \tag{4.6}$$

where t_i is a standard normal variable. Using the estimates coefficients $(\hat{\alpha})$, willingness to pay can be derived as[21]

$$E(\log(\mathrm{WTP})) = x'\hat{\alpha}, \tag{4.7a}$$

[20]True values are as theoretically defined in the examples provided in Eqs. (4.1) and (4.2).

[21]Wooldridge (2012, pp. 212-213) noted that the log-normal estimator is consistent but could be biased.

or

$$E(\text{WTP}) = \exp(x'\hat{\alpha})\exp\left(\frac{\hat{\sigma}^2}{2}\right). \tag{4.7b}$$

The point estimate of WTP that results from Eq. (4.7b) is a function of chosen levels of the x vector, and several computation approaches are generally followed. The first is to insert mean values for each element of x to predict WTP. The second is to insert the unique values for the elements of x for each person in the sample, predict the individual-specific WTP, and then compute mean WTP from the predictions. To compute WTP for a specific change in quality (or quantity), the variable(s) in x representing the change need to be set to the appropriate level(s).

Again, following Cameron and Huppert (1989)[22] and continuing with the notation established for the analysis of payment-card data, analysis of dichotomous-choice data proceeds as follows:

$$\begin{aligned}\Pr(\text{yes}_i) &= \Pr(\log(\text{WTP}_i) > \log \$\text{B}_i) \\ &= \Pr\left(\frac{e_i}{\sigma} > \frac{\log \$\text{B}_i - x'\alpha}{\sigma}\right) \\ &= 1 - \emptyset\left(\frac{\log \$\text{B}_i - x'\alpha}{\sigma}\right).\end{aligned} \tag{4.8}$$

The computation of mean and median values proceeds as described above for payment-card data.

The issue of a spike in the probability distribution at \$0 for people who do not value the change has received limited attention in the literature. In order to address zero values in the analysis of contingent-valuation data, this issue needs to be considered early in the study design to identify people who hold values of \$0. With this information, open-ended responses can be analyzed using a Tobit model (Jones et al. 2008; Strazzera et al. 2003), and alternative specifications have been considered for payment-card and dichotomous-choice data (Kriström 1997; Bohara et al. 2001).

In addition to the parametric approach described above, semiparametric and nonparametric approaches have been used to analyze dichotomous-choice data (Araña and León 2005; Creel and Loomis 1997; Crooker and Herriges 2004; Fernández et al. 2004; Haab and McConnell 1997; Kriström 1990; Li 1996). A nonparametric approach reduces the distribution and functional form assumptions that must be imposed on the data when estimating WTP. A lower-bound nonparametric estimator is often used to compute WTP and this estimator is specified as

[22]Hanemann (1984) proposed an alternative random-utility approach to analyzing dichotomous-choice data, and McConnell (1990) showed the Hanemann (1984), and Cameron and Huppert (1989) approaches to be the duals of each other.

$$\overline{\mathrm{WTP}_l} = \sum_{i=1}^{k} (\$\mathrm{B}_i - \$\mathrm{B}_{i-1}) * p(\$\mathrm{B}_i), \tag{4.9}$$

where $\overline{\mathrm{WTP}_l}$ is a lower-bound estimator of WTP, k is the number of bids used in the study design, and $p(\bullet)$ is the percentage of respondents that answered "yes" to bid amount $\$\mathrm{B}_i$. Here it is assumed that $\$\mathrm{B}_{i-1} = 0$ for $i = 1$.[23] This is a lower-bound estimator because anyone who might hold a value greater than the highest bid amount is assigned a value of \$0, and the percentage of people who answered "yes" in each bid interval are assigned the lower percentage of the higher bid. For discussion of the econometric analysis of this nonparametric lower-bound estimator, see Lewbel et al. (2011) and Wantanabe (2010).

It is also important to consider the variability of estimated WTP for studies to support decision-making and for experiments. Variance can be a function of heterogeneity in values across the sample and can be affected by the study design. That is, not everyone in a sample holds the same value, and larger within-sample variability will lead to larger standard errors of the estimated mean. In addition, the quality of the contingent-valuation scenario (and survey) can influence the estimated standard error; a poorer scenario design might result in a larger standard error of estimated WTP than a better scenario design. This can arise because there is less or greater understanding of the description of the item being valued. Considering variability in estimated willingness to pay is important to support decision-making and for experiments. Larger variances can lead to failure to reject the null hypotheses when the hypotheses should be rejected in statistical tests, and they can reduce the credibility of WTP estimates to support decision-making. Thus, it is important to compute the standard error of estimated WTP and consider design approaches that minimize undue variability in valuation responses through careful pretesting of valuation scenarios.

A number of approaches have been used to develop confidence intervals for WTP estimates (Cooper 1994). A quick review of recent articles suggests that two approaches are used the most commonly. The first is the well-known Krinsky-Robb approach that Park et al. (1991) introduced to the nonmarket valuation literature. The second is the convolutions approach introduced by Poe et al. (1994). See Cooper (1994), and Poe et al. (2005) for comparisons of the alternative approaches.

There are other approaches that have been used to analyze payment-card and dichotomous-choice response data, and I have simply presented the approachs here for illustrative purposes. Readers seeking more information on the econometric analysis of contingent-valuation data should see Haab and McConnell (2002).

[23]Haab and McConnell (1997) did not discuss this normalization in their presentation of the nonparametric estimator.

4.1.10 Report Study

The singular most important element in reporting a contingent-valuation study that supports decision-making is to present an aggregate welfare estimate (aggWTP). This calculation, in simplest terms, is

$$\mathrm{aggWTP} = \overline{\mathrm{WTP}} * N, \tag{4.10}$$

where $\overline{\mathrm{WTP}}$ is the estimated mean willingness to pay from the sample and N is the size of the affected population. But this calculation is not simple at all, and a number of questions need to be addressed:

- Will mean or median willingness to pay be used?
- Will weighting be used to bring respondent sample characteristics in line with the affected population characteristics?
- Should individuals who are affected by the change valued and eligible to participate in the study, but are precluded for some reason (e.g., a gated community not allowing access to interviewers) be included in N?
- How should people who resfuse to complete the survey or answer the valuation question be treated?

There are a variety of other questions in this line that might need to be addressed, and addressing these issues is critically important because they can substantially affect the magnitude of the aggregate value estimate.

Reporting of study results (Step 10) requires an understanding that the findings serve multiple purposes—the current decision analysis, transfers to new decision applications at the same application site or to an entirely new application, and advancing the literature by helping future investigations understand the study design, estimation, and testing. Even methodological studies, while not designed to support decision-making, could still be used in benefit transfers (Chap. 11). Thus, clear and detailed study documentation within the limits of the publication outlet is crucial for study credibility and advancing the contingent valuation literature. The steps (Table 4.1) in this chapter provide a guide to assist in this documentation, such as

- The study application.
- The theoretical definition of the estimated value that includes the current condition or baseline and the change valued.
- The steps in the design process and discussions of why items in the final design worked well, what design items were tried and discarded, and why they were discarded.
- The sample frame.

- Survey mode and response rates reported according to the American Association for Public Opinion Research guidelines (www.aapor.org/Standards-Ethics/Best-Practices.aspx).
- The verbatim commodity description and valuation scenario from the survey.
- The contingent-valuation format used, including the question wording.
- Respondents' demographic characteristics, and use or preferences for the item valued.
- Methods of data analysis, including treatment of $0 values and protest responses, and econometric analysis.
- Estimates of central tendency and dispersion, methods used to calculate these sample statistics, and any robustness checks.

This information allows readers to evaluate the content validity of value estimates for the current application and to evaluate the transferability of value estimates to new applications. With limitations on what can be reported in journal articles, best practices should include providing online appendices that fully document studies.

4.2 Reliability and Validity

A reliability study investigates the variance in value estimates. Tests of validity ask whether a contingent-valuation study accurately measures the value concept it is designed to estimate (see Chap. 12). The consideration of reliability and validity collectively constitute what is referred to as a credible or accurate contingent-valuation study.

4.2.1 Reliability

Reliability investigates the variance in value estimates. The common investigation approach is test–retest reliability, where a contingent-valuation survey is repeated at two different points in time. This can be a replication with the same respondents or a between-subjects design where the two samples are drawn identically from the same sample frame.

The consensus in the literature appears to support a conclusion that contingent-valuation estimates are reliable (Brown et al. 2008; Carson et al. 1997; Kealy et al. 1988, 1990; Loomis 1989, 1990; Onwujekwe and Fox-Rushby 2005; Reiling et al. 1990; Stevens et al. 1994; Teisl et al. 1995). Thus, reliability of contingent-valuation estimates is not an issue of concern.

It is also important to recognize that values can and should change over time. Thus, failure to establish statistical equivalence in values over time does not refute reliability in instances where values have legitimately changed.

4.2.2 Validity

Three types of validity are commonly investigated: content, construct and criterion (Carmines and Zeller 1979). Criterion validity compares contingent-valuation estimates to a measurement that is external to the contingent-valuation study and is a presumed measure of the true value. The cash transactions in the Bishop and Heberlein (1979) study provided such a criterion upon which the parallel contingent-valuation estimate is validated. Convergent validity, which is a specific type of construct validity, investigates the consistency of contingent-valuation estimates with estimates provided by another nonmarket valuation method. This is what was done when Bishop and Heberlein compared the contingent-valuation estimate with those derived from a travel-cost model. Content validity asks whether the elements in the design of the contingent-valuation survey and data analyses are consistent with economic theory, established practice, and the valuation objective.

Validity assessments are only as good as the counterfactual that they are compared against. The criterion must be a true representation of the economic construct being valued. Convergent validity only establishes whether two estimation approaches are comparable; comparability can hold when both approaches have the same bias. Or, failure to establish convergent validity only implies that one or both estimates are biased. Two types of criterion validity studies are discussed: comparisons with cash experiments and comparisons with outcomes of actual referendum votes. Content validity requires a collective protocol for what constitutes best practices; the material provided in this chapter is one such documentation of best practices.

Researchers have conducted meta-analyses of studies using experiments to investigate differences in outcomes between contingent-valuation (or stated preference) treatments and control treatments where a simulated market is used that involves cash transactions (List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005). List and Gallet reviewed 29 studies that provided 174 validity comparisons.[24] Note, the comparisons are not all based on contingent-valuation studies, but experiments where stated preference value estimates are compared to "parallel" values estimated using cash transactions, and the comparisons may be for private goods and not public good applications. List and Gallet report a calibration factor for each study, which is the stated-preference value divided by the cash value. They conclude that the "calibration factors for the most prevalent type of study are 1.26 (minimum), 1.28 (median), and 1.30 (maximum), which suggests that the most common type of (WTP) study will tend to produce a slightly (upward) biased estimate of the actual value" (p. 250). This difference (calibration factor) is what is known as hypothetical bias. However, the credence of this conclusion crucially depends on two features: first, the experiment is similar to what would be conducted

[24]Little and Berrens (2004), and Murphy et al. (2005) analyzed essentially the same data. Little and Berrens only considered the percentage choosing to buy and not the welfare estimates, and Murphy et al. is a reanalysis of List and Gallet (2001).

in a contingent-valuation experiment, and second, the cash value measures the same underlying Hicksian concept of value as contingent valuation.

Some researchers have compared contingent-valuation study results to those from a parallel referendum vote. While many have asked how contingent-valuation estimates would compare to values if a market existed, comparisons to referendum votes make logical sense. Many contingent-valuation applications value public goods where a majority vote would be a logical decision rule. Further, the preferred question framing is as a dichotomous choice on a referendum vote. Vossler and Kerkvliet (2003) found that "survey responses match the actual voting outcome and WTP estimates based on the two (survey and voting) are not statistically different" (p. 631). Johnston (2006) similarly found no statistical difference between the proportion of "yes" votes in a stated preference study and an actual referendum at the expected program cost per household. Vossler et al. (2003) found a similar result when undecided votes were excluded or treated as "no" responses in the contingent-valuation portion of the experiment. These results indicate that responses to contingent-valuation questions, framed as votes to a referendum, mimic how people vote when they go to polls to vote on environmental issues. While there have been far fewer of these studies than of the comparisons with cash transactions discussed above, these studies indicate that comparing contingent-valuation outcomes with referendum votes is an important and promising line of inquiry.[25]

Finally, the discussion of validity has been extended to investigations of whether contingent-valuation results follow an expected pattern, which has been termed a "test of scope." A test of scope asks the simple question, "Are people willing to pay more for a larger change in the item being valued than for a smaller change?" but the conduct of these tests is varied (Desvousges et al. 1993). A meta-analysis of scope experiments suggests that contingent-valuation studies do pass a scope test (Ojea and Loureiro 2011).

Some, however, have argued that a scope test should address more than the simple question of whether more is valued more highly than less of the item being valued (e.g., the adding-up test; Diamond 1996; Desvousges et al. 2012). However, for these additional tests to be applied, assumptions about the structure of preferences must be imposed (Haab et al. 2013). Thus, if an empirical test fails, it is impossible to differentiate failure due to a problem in the structure of a contingent-valuation study or failure due to the wrong assumptions about preferences being imposed. Thus, while a scope test can provide insight about the credibility of a contingent-valuation estimate of value, it is a weak test of validity at best (see Heberlein et al. 2005).

More will be said about these points and other considerations of validity in Chap. 12.

[25]Schläpfer et al. (2004) also compared contingent-valuation results with voting behavior and found differences that indicate contingent valuation leads to overestimates of value, but confounds in this study limit the generalizability of the result.

4.2.3 Enhancing Validity

Based on the presumption that contingent valuation leads to overestimates of what people would actually pay (List and Gallet 2001), a number of protocols have been proposed to remove overestimation. These include the lower-bound nonparametric estimator of WTP discussed above (Haab and McConnell 1997), insertion of cheap talk into a survey instrument (Cummings and Taylor 1999), use of self-reported uncertainty scales after respondents have answered the contingent-valuation question (Champ et al. 1997), and consequentiality (Carson and Groves 2007). The latter of these approaches holds the most promise.

Each of these approaches leaves questions to be answered. Is it desirable to provide a lower-bound estimate of WTP that might actually underestimate true willingness to pay? Does cheap talk provide the correct incentive to provide truthful answers to a contingent-valuation question or just challenge people to lower their value responses? What is the basis for the correct interpretation of the uncertainty scale, and does it vary over specific valuation applications? Consequentiality, which requires a binding payment and a non-zero probability that the value estimates will inflence decision making, holds the most promise (Carson et al. 2014; Vossler et al. 2012).

4.3 Conclusions

It is important to recognize that the design of a contingent-valuation study involves constraints and trade-offs. The constraints could come from the actual change being valued, where the realism challenges the design (e.g., a tax will be used to fund the provision of the item to be valued, but pretesting reveals people value the item and oppose a tax increase). Trade-offs could arise in an attempt to develop a clear scenario while also seeking brevity. These decisions are part of the "art" of contingent valuation and have provided opportunities for critics to question the credibility of contingent-valuation estimates. This criticism is not all bad as it has led to better-designed studies and more-focused validity research. The outcome has been improved contingent-valuation studies that are in common use to support public and private decision-making.

It is important also to realize that the influence the survey design has over the outcome of any contingent-valuation study is **not** any different from any other line of research or empirical analysis. Simply put, contingent valuation—like any other empirical method—requires a high degree of skill and considerable art that must be combined with careful pretesting and validity checks.

Finally, the next chapter (Chap. 5) discusses choice experiments, which are a close cousin to contingent valuation in the family of stated preference valuation methods. There are some unique differences between these valuation approaches:

- Contingent-valuation scenarios define the change to be valued in a written scenario, while choice-experiment scenarios define the change using specific levels of attributes.
- Contingent valuation, using a dichotomous-choice question, asks respondents to choose between the change and a status quo condition, while choice experiments typically ask respondents to choose between two or more changes (alternatives) and a status quo condition.
- Contingent-valuation studies often include one valuation question, while choice experiments typically include multiple valuation questions.

The key distinction between contingent valuation and choice experiments is the scenario presentation; in a choice experiment respondents are presented with alternatives to choose among (usually three or more) that are described in terms of attributes (again, usually three or more). Whether contingent valuation or a choice experiment should be employed is not clear-cut. However, if multiple alternatives are not plausible and the item being valued cannot be logically divided into attributes from a design, policy, or consumer-preference perspective, contingent valuation is the appropriate choice. If multiple alternatives and attributes are plausible, the policy question seeks marginal values for individual attributes and such segmentation into attributes is plausible to respondents, a choice experiment would be the appropriate method. A choice experiment with one alternative, two attributes (including cost), and one valuation question only varies from a dichotomous-choice question in the presentation of the change to be valued, and they are conceptually and analytically equivalent. Thus, many of the design insights discussed in this chapter also relate to choice experiments and vice versa for much of the design feature discussion in Chap. 5.

Acknowledgments This chapter benefits from comments by other authors in this book at the USDA Forest Service sponsored workshop in Estes Park, Colorado. Review comments by Vic Adamowicz, Rich Bishop, Patty Champ, Kerry Smith, and Weizhe Weng are greatly appreciated. All errors and omissions are solely my responsibility.

References

Adamowicz, W., Dickie, M., Gerking, S., Veronesi, M. & Zinner, D. (2014). Household decision making and valuation of environmental health risks to parents and their children. Journal of the Association of Environmental and Resource Economists, 1, 481-519.

Ahearn, M. C., Boyle, K. J. & Hellerstein, D. R. (2003). Designing a contingent-valuation study to estimate the benefits of the conservation reserve program on grassland bird populations. In J. Kahn, D. Bjornstad & A. Alberini (Eds.), The contingent-valuation handbook. Cheltenham, United Kingdom: Edward Elgar.

Ajzen, I., Brown, T. C. & Rosenthal, L. H. (1996). Information bias in contingent valuation: Effects of personal relevance, quality of information, and motivational orientation. Journal of Environmental Economics and Management, 30, 43-57.

Alberini, A. (1995a). Optimal designs for discrete choice contingent valuation surveys: Single-bound, double-bound, and bivariate models. Journal of Environmental Economics and Management, 28, 287-306.
Alberini, A. (1995b). Willingness-to-pay models of discrete choice contingent valuation survey data. Land Economics, 71, 83-95.
Alberini, A., Boyle, K. & Welsh, M. (2003). Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management, 45, 40-62.
Alberini, A., Rosato, P., Longo, A. & Zanatta, V. (2005). Information and willingness to pay in a contingent valuation study: The value of S. Erasmo in the lagoon of Venice. Journal of Environmental Planning and Management, 48, 155-175.
Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Richardian equivalence. Journal of Political Economy, 97, 1447-1458.
Araña, J. E. & León, C. J. (2005). Flexible mixture distribution modeling of dichotomous choice contingent valuation with heterogeneity. Journal of Environmental Economics and Management, 50, 170-188.
Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. & Schuman, H. (1993). Natural resource damage assessments under the Oil Pollution Act of 1990. Federal Register, 58, 4601-4614.
Bateman, I. J., Day, B. H., Dupont, D. P. & Georgiou, S. (2009). Procedural invariance testing of the one-and-one-half-bound dichotomous choice elicitation method. The Review of Economics and Statistics, 91, 806-820.
Bateman, I. J., Day, B. H., Georgiou, S. & Lake, I. (2006). The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP. Ecological Economics, 60, 450-460.
Bateman, I. J., Jones, A. P., Lovett, A. A., Lake, I. R. & Day, B. H. (2002). Applying geographic information systems (GIS) to environmental and resource economics. Environmental and Resource Economics, 22, 219-269.
Bateman, I. J., Langford, I. H., Jones, A. P. & Kerr, G. N. (2001). Bound and path effects in double and triple bounded dichotomous choice contingent valuation. Resource and Energy Economics, 23, 191-213.
Bateman, I. J., Langford, I. H., Turner, R. K., Willis, K. G. & Garrod, G. D. (1995). Elicitation and truncation effects in contingent valuation studies. Ecological Economics, 12, 161-180.
Bateman, I. J. & Munro, A. (2009). Household versus individual valuation: What's the difference? Environmental and Resource Economics, 43, 119-135.
Becker, G. S. (1981). Treatise on the family. Cambridge, MA: Howard University Press.
Belsey, D. A., Kuh, E. & Welsch, R. E. (1980). Regression diagnostics. New York: Wiley.
Bergstrom, J. C., Stoll, J. R. & Randall, A. (1990). The impact of information on environmental commodity valuation decisions. American Journal of Agricultural Economics, 72, 614-621.
Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L. & Weimer, D. L. (2004). Information and effort in contingent valuation surveys: Application to global climate change using national internet samples. Journal of Environmental Economics and Management, 47, 331-363.
Berrens, R. P., Bohara, A. K., Jenkins-Smith, H., Silva, C. L., Ganderton, P. & Brookshire, D. (1998). Joint investigation of public support and public values: Case of instream flows in New Mexico. Ecological Economics, 27, 189-203.
Berrens, R. P., Brookshire, D., Ganderton, P. & McKee, M. (1998). Exploring nonmarket values for the social impacts of environmental policy change. Resource and Energy Economics, 20, 117-137.
Bishop, R. C. (1982). Option value: An exposition and extension. Land Economics, 58, 1-15.
Bishop, R. C. & Heberlein, T. A. (1979). Measuring values of extra-market goods: Are indirect measures biased? American Journal of Agricultural Economics, 61, 926-930.
Blamey, R. K., Bennett, J. W. & Morrison, M. D. (1999). Yea-saying in contingent valuation surveys. Land Economics, 75, 126-141.

Bohara, A. K., Kerkvliet, J. & Berrens, R. P. (2001). Addressing negative willingness to pay in dichotomous choice contingent valuation: A Monte Carlo simulation. Environmental and Resource Economics, 20, 173-195.

Bohara, A. K., McKee, M., Berrens, R. P., Jenkins-Smith, H., Silva, C. L. & Brookshire, D. S. (1998). Effect of total cost and group-size information on willingness to pay responses: Open ended vs. dichotomous choice. Journal of Environmental Economics and Management, 35, 142-163.

Boyle, K. J. (1989). Commodity specification and the framing of contingent-valuation questions. Land Economics, 65, 57-63.

Boyle, K. J. & Bishop, R. C. (1988). Welfare measurements using contingent valuation: A comparison of techniques. American Journal of Agricultural Economics, 70, 20-28.

Boyle, K. J., Bishop, R. C. & Welsh, M. P. (1985). Starting point bias in contingent valuation bidding games. Land Economics, 61,188-196.

Boyle, K. J., Desvousges, W. H., Johnson, F. R., Dunford, R. W. & Hudson, S. P. (1994). An investigation of part-whole biases in contingent-valuation studies. Journal of Environmental Economics and Management, 27, 64-83.

Boyle, K. J., Johnson, F. R. & McCollum, D. W. (1997). Anchoring and adjustment in single-bounded, contingent-valuation questions. American Journal of Agricultural Economics, 79, 1495-1500.

Boyle, K. J., MacDonald, H. F., Cheng, H. & McCollum, D. W. (1998). Bid design and yea saying in single-bounded, dichotomous-choice questions. Land Economics, 74, 49-64.

Boyle, K. J., Morrison, M., MacDonald, D. H., Duncan, R. & Rose, J. (2016). Investigating Internet and mail implementation of stated-preference surveys while controlling for differences in sample frames. Environmental and Resource Economics, 64, 401–419.

Brown, T. C., Kingsley, D., Peterson, G. L., Flores, N. E., Clarke, A. & Birjulin, A. (2008). Reliability of individual valuations of public environmental goods: Choice consistency, response time, and preference refinement. Journal of Public Economics, 92, 1595-1606.

Cameron, T. A. & Huppert, D. D. (1989). OLS versus ML estimation of non-market resource values with payment card interval data. Journal of Environmental Economics and Management, 17, 230-246.

Cameron, T. A., Shaw, W. D. & Ragland, S. R. (1999). Nonresponse bias in Maine survey data: Salience vs. endogenous survey complexity. In J. A. Herriges & K. L. Kling (Eds.), Valuing recreation and the environment: Revealed preference methods in theory and practice (pp. 217–251). Cheltenham, United Kingdom: Edward Elgar.

Carmines, E. G. & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills, CA: Sage.

Campos, P., Caparrós, A. & Oviedo, J. L. (2007). Comparing payment-vehicle effects in contingent valuation studies for recreational use in two protected Spanish forests. Journal of Leisure Research, 39, 60-85.

Carson, R. T. & Groves, T. (2007). Incentive and informational properties of preference questions. Environmental and Resource Economics, 37, 181-210.

Carson, R. T., Groves, T. & List, J. A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. Journal of the Association of Environmental and Resource Economists, 1, 171-207.

Carson, R. T., Hanemann, W. M., Kopp, R., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A. & Smith, V. K. (1998). Referendum design and contingent valuation: The NOAA panel's no-vote recommendation. Review of Economics and Statistics, 80 (3), 484-487.

Carson, R. T., Hanemann, W. M., Kopp, R., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Conaway, M. & Martin, K. (1997). Temporal reliability of estimates from contingent valuation. Land Economics, 73, 151-163.

Champ, P. A., Bishop, R. C., Brown, T. C. & McCollum, D. W. (1997). Using donation mechanisms to value nonuse benefits from public goods. Journal of Environmental Economics and Management, 33, 151-162.

Chiappori, P. A. & Ekeland, I. (2009). The microeconomics of efficient group behavior: Identification. Econometrica 67, 763-799.

Cicchetti, C. J. & Smith, V. K. (1973). Congestion, quality deterioration, and optimal use: Wilderness recreation in the Spanish peaks primitive area. Social Science Research, 2, 15-30.

Cooper, J. C. (1994). A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous-choice contingent valuation surveys. Land Economics, 70, 111-122.

Cooper, J. C., Hanemann, M. & Signorello, G. (2002). One-and-one-half-bound dichotomous choice contingent valuation. The Review of Economics and Statistics, 84, 742-750.

Covey, J., Loomes, G. & Bateman, I. J. (2007). Valuing risk reductions: Testing for range biases in payment card and random card sorting methods. Journal of Environmental Planning and Management, 50, 467-482.

Creel, M. & Loomis, J. (1997). Semi-nonparametric distribution free dichotomous choice CV. Journal of Environmental Economics and Management, 32, 341-358.

Crooker, J. R. & Herriges, J. A. (2004). Parametric and semi-nonparametric estimation of willingness-to-pay in the dichotomous choice contingent valuation framework. Environmental and Resource Economics, 27, 451-480.

Cummings, R. G., Brookshire, D. S. & Schulze, W. D. (Eds.). (1986). Valuing environmental goods: An assessment of the contingent valuation method. Totowa, NJ: Rowman & Allanheld.

Cummings, R. G. & Taylor, L. O. (1998). Does realism matter in contingent valuation surveys? Land Economics, 74, 203-215.

Cummings, R. G. & Taylor, L. O. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. American Economic Review, 89, 649-665.

Dalmau-Matarrodona, E. (2001). Alternative approaches to obtain optimal bid values in contingent valuation studies and to model protest zeros: Estimating the determinants of individuals' willingness to pay for home care services in day case surgery. Health Economics, 10, 101-118.

Davis, R. K. (1963). Recreation planning as an economic problem. Natural Resources Journal, 3, 239-249.

Davis, D. D. & Holt, C. A. (1993). Experimental economics. Princeton, NJ: Princeton University Press.

Desaigues, B., Ami, D., Bartczak, A., Braun-Kohlová, M., Chilton, S., Czajkowski, M., Farreras, V., Hunt, A., Hutchinson, M., Jeanrenaud, C., Kaderjak, P., Máca, V., Markiewicz, O., Markowska, A., Metcalf, H., Navrud, S., Nielsen, J. S., Ortiz, R., Pellegrini, S., Rabl, A., Riera, R., Scasny, M., Stoedckel, M. E., Szánto, R. & Urban, J. (2011). Economics valuation of air pollution mortality: A 9-country contingent valuation survey of value of a life year (VOLY). Ecological Indicators, 11, 902-910.

Desvousges, W. H., Johnson, F. R., Dunford, R. W., Hudson, S. P., Wilson, K. N. & Boyle, K. J. (1993). Measuring natural resource damages with contingent valuation. In J. A. Hausman (Ed.), Contingent valuation: A critical assessment (Contributions to economic analysis, vol. 220, pp. 91–164). Amsterdam, The Netherlands: Emerald Group.

Desvousges, W. H., Mathews, H. K. & Train, K. (2012). Adequate responsiveness to scope in contingent valuation. Ecological Economics, 84, 121-128.

Desvousges, W. H., Smith, V. K. & Fisher, A. (1987). Option price estimates for water quality improvements: A contingent valuation study for the Monongahela River. Journal of Environmental Economics and Management, 14, 248-267.

Diamond, P. (1996). Testing the internal consistency of contingent valuation surveys. Journal of Environmental Economics and Management, 30, 337-347.

Dillman, D. A. (2007). Mail and Internet surveys: The tailored design method. Hoboken, NJ: John Wiley & Sons.

Donaldson, C., Thomas, R. & Torgerson, D. J. (1997). Validity of open-ended and payment scale approaches to eliciting willingness to pay. Applied Economics, 29, 79-84.

Dupont, D. P. (2004). Do children matter? An examination of gender differences in environmental valuation. Ecological Economics, 49, 273-286.

Farreras, V., Riera, P. & Mogas, J. (2005). Does gender matter in valuation studies? Evidence from three forestry applications. Forestry, 78, 239-248.

Fernández, C., León, C. J., Steel, M. F. J. & Vázquez-Polo, F. J. (2004). Bayesian analysis of interval data contingent valuation models and pricing policies. Journal of Business Economics and Statistics, 22, 431-442.

Flores, N. E. & Carson, R. T. (1997). The relationship between the income elasticities of demand and willingness to pay. Journal of Environmental Economics and Management, 33, 287-295.

Foster, W. & Just, R. E. (1989). Measuring welfare effects of product contamination with consumer uncertainty. Journal of Environmental Economics and Management, 17, 266-283.

Freeman, A. M., III. (1993). The measurement of environmental and resource values: Theory and methods. Washington, DC: Resources for the Future.

Gabor, A. & Granger, C. W. J. (1966). Price as an indicator of quality. Economica, 33, 43-70.

Garcia-Llorente, M., Martin-López, B. & Montes, C. (2011). Exploring the motivations of protestors in contingent valuation: Insights for conservation policies. Environmental Science & Policy, 14, 76-88.

Green, D., Jacowitz, K., Kahneman, D. & McFadden, D. (1998). Referendum contingent valuation, anchoring and willingness to pay for public goods. Resource and Energy Economics, 20, 85-116.

Greenley, D. A., Walsh, R. G. & Young, R. A. (1981). Option value: Empirical evidence from a case study of recreation and water quality. Quarterly Journal of Economics, 96, 657-672.

Groothuis, P. A. & Whitehead, J. C. (2002). Does don't know mean no? Analysis of 'don't know' responses in dichotomous choice contingent valuation questions. Applied Economics, 34, 1935-1940.

Groves, R. M., Brick, J. M., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Krueter, F., Pennell, B.-E., Raghunathan, T., Schouten, B., Smith, T., Tourangeau, R., Bowers, A., Jans, M., Kennedy, C., Levenstein, R., Olson, K., Peytcheve, E., Ziniel, S. & Wager, J. (2008). Issues facing the field: Alternative practical measures of representativeness of survey respondent pools. Survey Practice, 1 (3), 1-6.

Groves, R. M. & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly, 72, 167-189.

Haab, T. C., Interis, M. G., Petrolia, D. R. & Whitehead, J. C. (2013). From hopeless to curious? Thoughts on Hausman's 'dubious to hopeless' critique of contingent valuation. Applied Economic Perspectives and Policy, 35, 593-612.

Haab, T. C. & McConnell, K. E. (1997). Referendum models and negative willingness to pay: Alternative solutions. Journal of Environmental Economics and Management, 32, 251-270.

Haab, T. C. & McConnell, K. E. (2002). Valuing environmental and natural resources: The econometrics of non-market valuation. Cheltenham, United Kingdom: Edward Elgar.

Hammack, J. & Brown, G. M., Jr. (1974). Waterfowl and wetlands: Toward bioeconomic analysis. Baltimore, MD: Johns Hopkins University Press.

Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. American Journal of Agricultural Economics, 66, 332-341.

Hanemann, W. M., Loomis, J. & Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. American Journal of Agricultural Economics, 73, 1255-1263.

Hanley, N., Schläpfer, F. & Spurgeon, J. (2003). Aggregating the benefits of environmental improvements: Distance-decay functions for use and nonuse values. Journal of Environmental Management, 68, 297-304.

Harbaugh, W. T. (1998). What do donations buy? A model of philanthropy based on prestige and warm glow. Journal of Public Economics, 67, 269-284.

Hausman, J. A. (Ed.). (1993). Contingent valuation: A critical assessment. Amsterdam, The Netherlands: Elsevier Science.

Hausman, J. A. (2012). Contingent valuation: From dubious to hopeless. Journal of Economic Perspectives, 25 (4), 43-56.

Heberlein, T. A., Wilson, M. A., Bishop, R. C. & Schaeffer, N. C. (2005). Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management, 50, 1-22.

Herriges, J. A. & Shogren, J. F. (1996). Starting point bias in dichotomous choice valuation with follow-up questioning. Journal of Environmental Economics and Management, 30, 112-131.

Hoehn, J. P. & Loomis, J. B. (1993). Substitution effects in the valuation of multiple environmental programs. Journal of Environmental Economics and Management, 25, 56-75.

Hoehn, J. P. & Randall, A. (1987). A satisfactory benefit-cost indicator from contingent valuation. Journal of Environmental Economics and Management, 14, 226-247.

Johnston, R. J. (2006). Is hypothetical bias universal? Validating contingent valuation responses using a binding referendum. Journal of Environmental Economics and Management, 52, 469-481.

Johnston, R.J., Boyle, K.J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., and Vossler, C.A. (2017). Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists (forthcoming).

Johnston, R. J., Schultz, E. T., Segerson, K., Besedin, E. & Ramachandran, M. (2012). Enhancing the content validity of stated preference valuation: The structure and function of ecological indicators. Land Economics, 88, 102-120.

Johnston, R. J., Swallow, S. K. & Weaver, T. F. (1999). Estimating willingness to pay and resource tradeoffs with different payment mechanisms: An evaluation of a funding guarantee for watershed management. Journal of Environmental Economics and Management, 38, 97-120.

Jones, N., Sophoulis, C. M. & Malesios, C. (2008). Economic valuation of coastal water quality and protest responses: A case study in Mitilini, Greece. The Journal of Socio-Economics, 37, 2478-2491.

Kanninen, B. J. (1993a). Design of sequential experiments for contingent valuation studies. Journal of Environmental Economics and Management, 25, S1-S11.

Kanninen, B. J. (1993b). Optimal experimental design for double-bounded dichotomous choice contingent valuation. Land Economics, 69,138-146.

Kanninen, B. J. (1995). Bias in discrete response contingent valuation. Journal of Environmental Economics and Management, 28, 114-125.

Kealy, M. J., Dovidio, J. F. & Rockel, M. L. (1988). Accuracy in valuation is a matter of degree. Land Economics, 64, 158-171.

Kealy, M. J., Montgomery, M. & Dovidio, J. F. (1990). Reliability and predictive validity of contingent values: Does the nature of the good matter? Journal of Environmental Economics and Management, 19, 244-263.

Keeter, S., Miller, C., Kohut, A., Groves, R. M. & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. Public Opinion Quarterly, 64, 125-148.

Kling, C. L., Phaneuf, D. J. & Zhao, J. (2012). From Exxon to BP: Has some number become better than no number? Journal of Economic Perspectives, 25 (4), 3-26.

Kotchen, M. J. & Reiling, S. D. (1999). Do reminders of substitutes and budget constrains influence contingent valuation estimates? Another comment. Land Economics, 75, 478-482.

Kreuter, F., Presser, S. & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question sensitivity. Public Opinion Quarterly, 72, 847-865.

Kriström, B. (1990). A non-parametric approach to the estimation of welfare measures in discrete response valuation studies. Land Economics, 66, 135-139.

Kriström, B. (1997). Spike models in contingent valuation. American Journal of Agricultural Economics, 79, 1013-1023.

Leggett, C. G., Kleckner, N. S., Boyle, K. J., Duffield, J. W. & Mitchell, R. C. (2003). Social desirability bias in contingent valuation surveys administered through in-person interviews. Land Economics, 79, 561-575.

Lewbel, A., McFadden, D. & Linton, O. (2011). Estimating features of a distribution from binomial data. Journal of Econometrics, 162, 170-188.

Li, C. Z. (1996). Semiparametric estimation of the binary choice model for contingent valuation. Land Economics, 72, 462-473.

Li, X., Boyle, K. J., Holmes, T. P. & Pullis-LaRouche, G. (2014). The effect of on-site forest experience on stated preferences for low impact timber harvesting programs. Journal of Forest Economics, 20, 348-362.

Lindhjem, H. & Navrud, S. (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuations? Ecological Economics, 70, 1628-1637.

Lindhjem, H. & Navrud, S. (2012). Asking for individual or household willingness to pay for environmental goods? Environmental and Resource Economics, 43, 11-29.

List, J. A. & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? Evidence from a meta-analysis. Environmental and Resource Economics, 20, 241-254.

Little, J. & Berrens, R. (2004). Explaining disparities between actual and hypothetical stated values: Further investigation using meta-analysis. Economics Bulletin, 3 (6), 1-13.

Loomis, J. B. (1989). Test-retest reliability of the contingent valuation method: A comparison of general population and visitor responses. American Journal of Agricultural Economics, 71, 76-84.

Loomis, J. B. (1990). Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques. Journal of Environmental Economics and Management, 18, 78-85.

Loomis, J., Gonzalez-Caban, A. & Gregory, R. (1994). Do reminders of substitutes and budget constraints influence contingent valuation estimates? Land Economics, 70, 499-506.

MacMillan, D., Hanley, N. & Leinhoop, N. (2006). Contingent valuation: Environmental polling or preference engine? Ecological Economics, 60, 299-307.

Madureira, L., Nunes, L. C., Borges, J. G. & Falcão, A. O. (2011). Assessing forest management strategies using a contingent valuation approach and advanced visualisation techniques: A Portuguese case study. Journal of Forest Economics, 17, 399-414.

Maguire, K. B., Taylor, L. O. & Gurmu, S. (2003). Do students behave like adults? Evidence from valuation experiments. Applied Economics Letters, 10, 753-756.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I. & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. Journal of the Market Research Society, 50, 79.

Mansfield, C., Van Houtven, G., Henderschott, A., Chen, P., Potter, J., Nourani, V. & Kilambi, V. (2012). Klamath River basin restoration nonuse value survey. Final report to the U.S. Bureau of Reclamation, RTI International, RTI Project Number 0212485.001.010.

Marwell, G. & Ames, R. E. (1981). Economists free ride, does anyone else? Experiments on the provision of public goods. Journal of Public Economics, 15 (3), 295-310.

McConnell, K. E. (1990). Models for referendum data: The structure of discrete choice models for contingent valuation. Journal of Environmental Economics and Management, 18, 19-34.

McFadden, D. (1994). Contingent valuation and social choice. American Journal of Agricultural Economics, 76 (4), 689-708.

Menges, R. & Beyer, G. (2014). Underground cables versus overhead lines: Do cables increase social acceptance of grid development? Results of a contingent valuation survey in Germany. International Journal of Sustainable Energy Planning and Management, 3, 33-48.

Messonnier, M. L., Bergstrom, J. C., Cornwell, C. M., Teasley, R. J. & Cordell, H. K. (2000). Survey response-related biases in contingent valuation: Concepts, remedies, and empirical application to valuing aquatic plant management. American Journal of Agricultural Economics, 82 (2), 438-450.

Meyerhoff, J. & Liebe, U. (2006). Protest beliefs in contingent valuation: Explaining their motivation. Ecological Economics, 57 (4), 583-594.

Mitchell, R. C. & Carson, R. T. (1981). An experiment in determining willingness to pay for national water quality improvements. Unpublished report. Washington, DC: Resources for the Future.

Mitchell, R. C. & Carson, R. T. (1989). Using surveys to value public goods: The contingent valuation method. Washington, DC: Resources for the Future.

Mitchell, R. C. & Carson, R. T. (1993). The value of clean water: The public's willingness-to-pay for boatable, fishable, and swimmable quality water. Water Resources Research, 29 (7), 2445-2454.

Munro, A. (2005). Household willingness to pay equals individual willingness to pay if and only if the household income pools. Economics Letters, 88, 227-230.

Murphy, J. J., Stevens, T. H. & Yadav, L. (2010). A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. Environmental and Resource Economics, 47, 111-123.

Murphy, J. J., Allen, P. G., Stevens, T. H. & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics, 30, 313-325.

Nunes, P. A. & Schokkaert, E. (2003). Identifying the warm glow effect in contingent valuation. Journal of Environmental Economics and Management, 45, 231-245.

Ojea, E. & Loureiro, M. L. (2011). Identifying the scope effect on a meta-analysis of biodiversity valuation studies. Resource and Energy Economics, 33, 706-724.

Onwujekwe, O. & Fox-Rushby, J. (2005). Inter-rater and test-retest reliability of three contingent valuation question formats in south-east Nigeria. Health Economics, 14, 529-536.

Park, T., Loomis, J. & Creel, M. (1991). Confidence intervals for evaluating benefit estimates from dichotomous choice contingent valuation studies. Land Economics, 67, 64-73.

Poe, G. L. & Bishop, R. C. (1999). Valuing the incremental benefits of groundwater protection when exposure levels are known. Environmental and Resource Economics, 13, 347-373.

Poe, G. L. & Bishop, R. C. (2001). Information and the valuation of nitrates in ground water, Portage County, Wisconsin. In J. C. Bergstrom, K. J. Boyle & G. L. Poe (Eds.), The economic value of water quality. Cheltenham, United Kingdom: Edward Elgar.

Poe, G. L., Clark, J. E., Rondeau, D. & Schulze, W. D. (2002). Provision point mechanisms and field validity tests of contingent valuation. Environmental and Resource Economics, 23, 105-131.

Poe, G. L., Giraud, K. L. & Loomis, J. B. (2005). Computational methods for measuring the difference in empirical distributions. American Journal of Agricultural Economics, 87, 353-365.

Poe, G. L., Severance, E. K. & Welsh, M. P. (1994). Measuring the difference (X-Y) of simulated distributions: A convolutions approach. American Journal of Agricultural Economics, 76, 904-915.

Quiggin, J. (1998). Individual and household willingness to pay for public goods. American Journal of Agricultural Economics, 80, 58-63.

Ramajo-Hernández, J. & del Saz-Salazar, S. (2012). Estimating the non-market benefits of water quality improvement for a case study in Spain: A contingent valuation approach. Environmental Science & Policy, 22, 47-59.

Randall, A., Ives, B. & Eastman, C. (1974). Bidding games for evaluation of aesthetic environmental improvements. Journal of Environmental Economics and Management, 1, 132-149.

Reiling, S. D., Boyle, K. J., Phillips, M. L. & Anderson, M. W. (1990). Temporal reliability of contingent values. Land Economics, 66, 128-134.

Roach, B., Boyle, K. J. & Welsh, M. P. (2002). Testing bid design effects in multiple bounded contingent valuation. Land Economics, 78, 121-131.

Rowe, R. D., d'Arge, R. C. & Brookshire, D. S. (1980). An experiment on the economic value of visibility. Journal of Environmental Economics and Management, 7, 1-19.

Rowe, R. D., Schulze, W. D. & Breffle, W. S. (1996). A test for payment card biases. Journal of Environmental Economics and Management, 31, 178-185.

Ryan, M. & Watson, V. (2009). Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. Health Economics, 18, 389-401.

Salant, P. & Dillman, D. A. (1994). How to conduct your own survey. New York, NY: John Wiley & Sons.
Samples, K. C., Dixon, J. A. & Gowen, M. M. (1986). Information disclosure and endangered species valuation. Land Economics, 62, 306-312.
Scarpa, R. & Bateman, I. (2000). Efficiency gains afforded by improved bid design versus follow-up valuation questions in discrete-choice cv studies. Land Economics, 76, 299-311.
Schläpfer, F. (2006). Survey protocol and income effects in the contingent valuation of public goods: A meta-analysis. Ecological Economics, 57, 415-429.
Schläpfer, F., Roschewitz, A. & Hanley, N. (2004). Validation of stated preferences for public goods: A comparison of contingent valuation survey response and voting behavior. Ecological Economics, 51, 1-16.
Schneemann, M. (1997). A meta-analysis of response rates to contingent valuation surveys conducted by mail. Unpublished master's thesis. University of Maine.
Schouten, B., Cobben, F. & Bethlehem, J. (2009). Indicators for the representativeness of survey response. Survey Methodology, 35, 101-113.
Scott, A. (1965). The valuation of game resources: Some theoretical aspects. Canadian Fisheries Report, iv, Ottawa, Ontario, Canada: Department of Fisheries of Canada.
Shapiro, B. P. (1968). The psychology of pricing. Harvard Business Review, 46 (7), 14-25.
Smith, R. D. (2005). Sensitivity to scale in contingent valuation: The importance of the budget constraint. Journal of Health Economics, 24, 515-529.
Smith, V. L., Suchanek, G. L. & Williams, A. W. (1988). Bubbles, crashes and endogenous expectations in experimental spot asset markets. Econometrica, 56, 1119-1151.
Soliño, M., Vázquez, M. X. & Prada, A. (2009). Social demand for electricity from forest biomass in Spain: Does payment periodicity affect the willingness to pay? Energy Policy, 37, 531-540.
Stevens, T. H., DeCoteau, N. E. & Willis, C. E. (1997). Sensitivity of contingent valuation to alternative payment schedules. Land Economics, 73, 140-148.
Stevens, T. H., More, T. A. & Glass, R. J. (1994). Interpretation and temporal stability of CV bids for wildlife existence: A panel study. Land Economics, 70, 355-363.
Strazzera, E., Scarpa, R., Calia, P., Garrod, G. D. & Willis, K. G. (2003). Modeling zero values and protest responses in contingent valuation surveys. Applied Economics, 35, 133-138.
Teisl, M. F., Boyle, K. J., McCollum., D. W. & Reiling, S. D. (1995). Test-retest reliability of contingent valuation with independent sample pretestPretest and post-test control groups. American Journal of Agricultural Economics, 77, 613-619.
Teisl, M. F., Roe, B. & Hicks, R. D. (2002). Can eco-labels tuna market? Journal of Environmental Economics and Management, 43, 339-359.
Thayer, M. (1981). Contingent valuation techniques for assessing environmental impacts: Further evidence. Journal of Environmental Economics and Management, 8, 27-44.
Vossler, C. A., Doyon, M. & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. American Economic Journal: Microeconomics, 4 (4), 145-171.
Vossler, C. A. & Kerkvliet, J. (2003). A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum. Journal of Environmental Economics and Management, 45, 631-649.
Vossler, C. A., Kerkvliet, J., Polasky, S. & Gainutdinova, O. (2003). Externally validating contingent valuation: An open-space survey and referendum in Corvallis, Oregon. Journal of Economic Behavior & Organization, 51, 261-277.
Wantanabe, M. (2010). Nonparametric estimation of mean willingness to pay from discrete response valuation data. American Journal of Agricultural Economics, 92, 1114-1135.
Ward, K. L. & Duffield, J. W. (1992). Natural resource damages: Law and economics. New York: John Wiley & Sons.
Watson, V. & Ryan, M. (2007). Exploring preference anomalies in double bounded contingent valuation. Journal of Health Economics, 26, 463-482.

Welsh, M. P. & Poe, G. L. (1998). Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach. Journal of Environmental Economics and Management, 36, 170-185.

Whitehead, J. & Bloomquist, G. (1995). Do reminders of substitutes and budget constraints influence contingent valuation estimates? Comment. Land Economics, 71, 541-543.

Whittington, D. & Pagiola, S. (2012). Using contingent valuation in the design of payments for environmental services mechanisms: A review and assessment. World Bank Research Observer, 27, 261-287.

Wiser, R. H. (2007). Using contingent valuation to explore willingness to pay for renewable energy: A comparison of collective and voluntary payment vehicles. Ecological Economics, 62, 419-432.

Wooldridge, J. (2012). Introductory econometrics: A modern approach. Boston, MA: Cengage Learning.

Chapter 5
Choice Experiments

Thomas P. Holmes, Wiktor L. Adamowicz and Fredrik Carlsson

Abstract There has been an explosion of interest during the past two decades in a class of nonmarket stated-preference valuation methods known as choice experiments. The overall objective of a choice experiment is to estimate economic values for characteristics (or attributes) of an environmental good that is the subject of policy analysis, where the environmental good or service comprises several characteristics. Including price as a characteristic permits a multidimensional, preference-based valuation surface to be estimated for use in benefit-cost analysis or any other application of nonmarket valuation. The chapter begins with an overview of the historical antecedents contributing to the development of contemporary choice experiments, and then each of the steps required for conducting a choice experiment are described. This is followed by detailed information covering essential topics such as choosing and implementing experimental designs, interpreting standard and more advanced random utility models, and estimating measures of willingness-to-pay. Issues in implementing and interpreting random utility models are illustrated using a choice experiment application to a contemporary environmental problem. Overall, this chapter provides readers with practical guidance on how to design and analyze a choice experiment that provides credible value estimates to support decision-making.

Keywords Alternatives · Attributes · Choice set · Discrete-choice analysis · Experimental design · Nonmarket valuation · Passive-use value · Policy analysis · Questionnaire · Random utility model · Stated preference · Survey · Trade-offs · Use value · Welfare measures · Willingness to pay

T.P. Holmes (✉)
U.S. Forest Service, Southern Research Station, Research Triangle Park, NC, USA
e-mail: tholmes@fs.fed.us

W.L. Adamowicz
University of Alberta, Edmonton, AB, Canada
e-mail: vic.adamowicz@ualberta.ca

F. Carlsson
University of Gothenburg, Gothenburg, Sweden
e-mail: fredrik.carlsson@economics.gu.se

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_5

Stated-preference methods of environmental valuation where market transaction data have limitations or do not exist in a form useful for measurement of economic values have been used by economists for decades. There has been an explosion of interest during the past two decades in a class of stated-preference methods that we refer to as choice experiments (CEs).[1] The overall objective of a CE is to estimate economic values for characteristics (or attributes) of an environmental good that is the subject of policy analysis, where the environmental good or service comprises several characteristics. Including price as a characteristic permits a multidimensional, preference-based valuation surface to be estimated for use in benefit-cost analysis or any other application of nonmarket valuation.

CEs have gained popularity because they offer several potential advantages relative to other valuation methods.

- CEs can provide values for changes in a single characteristic or values for changes in levels of characteristics or values for multiple changes in characteristics, resulting in a response surface of values rather than a single value.
- Because characteristics are experimentally manipulated and presented to respondents, they are typically exogenous, not collinear, and can reflect characteristic levels outside the range of the current market or environment. This is in contrast with revealed preference data that are often collinear, may be limited in variation, and may be endogenous in explaining choices.
- Just like contingent valuation, CEs can be used to assess preferences or trade-offs in behavioral settings (e.g., recreation site choice) that are relevant for measuring use values or in settings that are used to measure passive-use values like voting or referenda.
- The CE presentation format makes choices relatively easy for respondents, and attributes and their levels can be customized such that they are realistic for respondents (i.e., reflecting specific conditions they face). These types of choices are often similar to those consumers face in markets.
- There is experience in using CEs in several disciplines, including marketing, transportation, and health economics. Environmental economists are typically more interested in welfare measures than are practitioners in other fields.
- The use of experimental design theory increases the statistical efficiency of the parameters estimated so that smaller samples may be used, which reduces implementation costs.

[1]The label “choice experiment” is a source of controversy. The previous edition of this book used the phrase “attribute-based methods” (which included ratings and rankings), while others have referred to this approach as “attribute-based stated choice methods,” “choice-based conjoint analysis,” and a host of other names. Carson and Louviere (2011) recommended the term “discrete choice experiment” to reflect the fact that these methods elicit a discrete response to an experimentally designed set of choice alternatives. Their definition includes what would normally be viewed as binary contingent valuation questions , as well as other variants of elicitation processes. This chapter focuses on what they refer to as a “multinomial choice sequence” (a series of multialternative experimentally designed choice questions).

Of course, these potential advantages come at a price, including the following challenges:

- As is the case in contingent valuation, CE responses are stated preferences, and concerns about strategic behavior or hypothetical bias arise.
- The cognitive difficulty faced by respondents in considering alternatives with multiple attributes in new choice situations may be high. Requiring respondents to assess complex trade-offs may result in behavioral responses such as the use of decision heuristics that are not well understood and might not reflect how they would make actual market choices.
- Experimental design theory is becoming more complex, and a sound grasp of the basic design principles is required to construct an experiment.
- The econometric models used to analyze CE data are becoming more complex and require advanced econometric and programming skills to operationalize.

However, while there are challenges, new software programs and tools have been developed that help in experimental design, data collection, and econometric analysis, and relatively straightforward procedures can be used to generate estimates that can be used in policy analysis.

In this chapter, the goal is to provide practical guidance on how to design and analyze a CE that provides credible value estimates to support decision-making. The chapter begins by describing the rich historical antecedents of modern applications of CEs (Sect. 5.1). This is followed by an overview of the basic steps required for conducting a choice experiment (Sect. 5.2). The next section expands on a set of selected topics in experimental design that are important to understand when developing a choice experiment (Sect. 5.3). Next, a description of the standard random utility model that provides the conceptual foundation for empirical analysis of CE data is presented (Sect. 5.4), along with an explanation of how to compute various measures of willingness to pay (Sect. 5.5). That is followed by descriptions of empirical models that relax standard assumptions, which are the subject of much current research (Sect. 5.6). To illustrate issues in implementation and interpretation of the standard and more advanced models, an application of a CE to an environmental problem is provided (Sect. 5.7). Concluding comments are presented (Sect. 5.8), followed by two examples illustrating recent applications of choice experiments (Appendices 1 and 2).

5.1 Interpretive History

The origins of CEs are found in various social science disciplines. Within economics, the conceptual foundation for CEs finds its source in the hedonic method that views the demand for goods as derived from the demand for attributes. This approach can be traced to Court (1939), who used hedonic regressions to study the demand for automobiles, and Griliches (1971), who used hedonic regressions in the construction of hedonic price indices. In the psychology literature, the comparative judgment approach (Thurstone 1927) and the judgment and decision-making literature (Hammond 1955; Anderson 1970) also include discussions of how consumers evaluate items and use these evaluations in choosing between items. Lancaster's (1966) theory of consumer demand provided the basic conceptual structure that underlies economic applications of CEs.

At the same time that Lancaster was writing about consumer demand being driven by commodity attributes, a new measurement technique in mathematical psychology was articulated for decomposing overall judgments regarding a set of complex alternatives into the sum of weights on attributes of the alternatives (Luce and Tukey 1964). The method was rapidly embraced by marketing researchers who recognized the value of information about the relative importance of commodity attributes in the design of new products (Green and Rao 1971; Green and Wind 1975). This new marketing research method became generally known as "conjoint analysis."

Many commercial applications for conjoint analysis were soon found, particularly the prediction of market share for new products (Cattin and Wittink 1982). The typical procedure would ask respondents to rate the attractiveness of a set of products and then model the preferences of each respondent.[2] This approach emphasized the importance of capturing individual-level preference heterogeneity as a key element in predicting market share.

Two primary concerns arose regarding the typical conjoint procedure. First, it was not clear that the information contained in rating data was the same as the information contained in choice data, which mimicked market transactions. Second, implementation of choice simulators, based on ratings data, was cumbersome and often confusing to managers who used the predictions of market share models.

A simpler, more direct approach to predicting choices in the marketplace was provided by discrete choice theory, particularly as formulated for econometric analysis by McFadden (1974). The conceptual foundation for McFadden's analysis of economic choice lay in Thurstone's (1927) idea of random utility. By positing that individuals make choices that maximize their utility and that not all determinants of choice are available for analysis, choice theory was placed on a strong economic foundation that included a richness of behavior not found in standard Hicks-Samuelson theory. In addition, starting with Luce's (1959) choice axiom, linked to the random utility model by Marschak (1960), McFadden developed an econometric model that combined hedonic analysis of alternatives and random utility maximization.[3] This model is known as the multinomial logit (conditional logit) model.

A further advance identified by McFadden and others is the linkage between random utility models and welfare economics. The utility function in random utility

[2]Rating scale approaches, or traditional conjoint analysis, are based on Torgerson's (1958) Law of Comparative Judgment. This approach presents individuals with profiles (alternatives) or bundles of attributes and asks them to provide a rating of each profile (e.g., 1 to 10, where 10 is very good, and 1 is very poor). The development of rating-based conjoint is discussed in Green and Srinivasan (1978) and Louviere (1988b).

[3]See also subsequent work by Manski (1977) and Yellot (1977).

models is actually a conditional indirect utility function (conditional on the choice of the alternative). Thus, including price as an attribute in the conditional indirect utility function allows one to assess economic welfare measures (Small and Rosen 1981).

The conceptual richness of random utility theory and the practical advantages of the multinomial logit model were embraced by marketing researchers who promoted the use of multinomial logit to analyze aggregate marketing data or choice data aggregated up to frequencies of choice (Louviere and Hensher 1983; Louviere and Woodworth 1983; Louviere 1988a). The random utility model also found wide application in transportation demand (Ben-Akiva and Lerman 1985; Louviere et al. 2000; Hensher et al. 2005). While initial work using the multinomial logit model was based on the analysis of aggregate or frequency of choice data, recent methodological developments have focused on understanding choices and sources of individual preference heterogeneity in random utility models, reminiscent of the focus on individual-level modeling used in conjoint analysis.

The first application of hedonic, stated-preference methods to environmental valuation was by Rae (1983), who used rankings (most preferred alternative, secondmost preferred, etc.) to value visibility impairments at U.S. national parks. Other environmental valuation studies using rankings include Smith and Desvousges (1986), who evaluated changes in water quality, and Lareau and Rae (1989), who estimated values for diesel odor reductions. Subsequent to these studies, ratings methods for environmental valuation, based on a Likert scale, grew in popularity (Gan and Luzar 1993; Mackenzie 1993; Roe et al. 1996). The popularity of ranking and rating methods for environmental valuation has diminished due to difficulties in linking such responses to responses consistent with economic theory (Louviere et al. 2010).

In the early 1990s, a number of applications of stated-preference experiments that used choices, rather than ratings or ranking, began to appear in the environmental economics literature. Among the first applications was Adamowicz et al. (1994), who demonstrated how revealed and stated-preference (choice experiment) data can be combined. Since then, the use of CEs in the literature has grown rapidly with applications to use values and passive-use or total values. At present, CEs are probably the most commonly used approach in the peer-reviewed literature within the class of discrete choice experiments. The literature on applications of CEs and on methodological issues surrounding CE implementation (particularly in the areas of experimental design and econometric analysis) has increased steadily, and "standard practice" changed dramatically over the last 30 years.

5.2 Steps in Conducting a Choice Experiment

Before deciding to conduct a choice experiment, it is essential to consider whether this method is the most appropriate or whether another technique, such as contingent valuation, would be better. The essence of this decision is whether it makes sense to frame a policy question in terms of the attributes and whether marginal values of the attributes are required for policy analysis. If a policy question, for example, seeks to identify forest management options that will provide the greatest benefit to moose hunters, then consumer choices between alternative moose hunting sites with different levels of attributes (such as moose abundance, road quality, and travel distance) provide a reasonable framework for analysis (Boxall et al. 1996). In contrast, if the policy question focuses on the value that hunters place on a moose hunting experience given current conditions, then a contingent valuation study may be a better approach (Boyle et al. 1996).

The second issue to consider is the technical composition of alternatives and the perception of attribute bundles by consumers. In the moose hunting example (Boxall et al. 1996), moose abundance, road quality, and travel distance can reasonably be considered to be independent attributes. This may not be the case for a suite of ecological characteristics that are technically linked in production (Boyd and Krupnik 2009).

If it is decided that a CE is the best approach for conducting policy analysis, then implementation should follow the seven steps outlined in Table 5.1 (based on Adamowicz et al. 1998). Each step is briefly described following the table.

5.2.1 Characterize the Decision Problem

The initial step in developing a CE is to clearly identify the dimensions of the problem. This requires thinking about two key issues: (1) the geographic and temporal scope of potential changes in policy attributes, and (2) the types of values that are associated with those changes. The geographic scope of a CE would include consideration of whose values are to be included in the valuation or benefit-cost analysis. If the value of a change in an endangered species management program is being considered, for example, should the CE be applied to people living in the region, province/state, country, or internationally? It is essential to identify who will be impacted by changes in policy attributes as well as to articulate how they will be impacted. In addition, if the policy context is specific to a geographic site, the location of substitute sites will be important in the design, as demonstrated in a tropical rainforest preservation study reported by Rolfe et al. (2000).

Table 5.1 Steps in implementing a choice experiment

Step 1	Characterize the decision problem
Step 2	Identify and describe the attributes
Step 3	Develop an experimental design
Step 4	Develop the questionnaire
Step 5	Collect data
Step 6	Estimate model
Step 7	Interpret results for policy analysis or decision support

Temporal considerations will also be important. There may be a need to include an attribute for program duration or when the benefits will accrue to the public (e.g., Qin et al. 2011).

The second issue is the type of value arising from the policy under consideration. Is the choice to be examined one that reflects use value or behavior (such as recreation site choice or choices of market goods), or is the choice best represented as a public choice (referendum) on a set of attributes arising from a policy change? The latter may contain both use and passive-use values—or it may reflect total economic value.

5.2.2 *Attribute Identification and Description*

Once the decision problem is characterized, it is necessary to identify and describe the relevant attributes, including the levels to be used for each attribute. Holding structured conversations (focus groups) with resource managers, scientists, and people who typify the population that will be sampled will help identify the important attributes. At this stage, it is often challenging to decide how many attributes to include in the experiment as well as the particular levels that each attribute can take.

Focus groups can be very useful in this case. Group members can be asked to describe what attributes they think of when considering the goods and services being affected by the policy. They can provide information on whether attributes and levels are credible, understandable, and clearly presented. Focus groups of policymakers and the public can be useful to identify whether the attributes being considered by policymakers coincide with those being evaluated by members of the public. However, focus groups will often provide long lists of attributes that could result in complex choice tasks. Because not much is known about how people respond to highly complex survey questions (Mazzotta and Opaluch 1995; Swait and Adamowicz 2001a, b), it is a good idea to keep the set of attributes and levels as simple as possible. Overall, focus groups are a very important and effective way to construct attributes, levels, and the appropriate framing of a choice task.

Describing attributes that represent passive-use values (such as the value of biodiversity conservation) can be particularly challenging. Boyd and Krupnick (2009) suggested that attributes should be thought of as endpoints that directly enter the utility functions or household production functions of consumers, or—if intermediate inputs are being considered—the pathway to the endpoint needs to be made clear. Thus, passive-use values associated with forest biodiversity, for example, can be described using indicators of species richness (Horne et al. 2005). However, because forest biodiversity can be influenced by forest management processes that are under the control of decision-makers, attributes could be described in terms of those processes so long as the linkages between processes and outcomes are made clear. Because individuals might be interested in the processes associated with the endpoint, it is important to clarify the things that people value, what decision-makers can affect, and the description of the attributes during this stage of survey development. In addition to identifying utility endpoints, Schultz et al. (2012) recommended further standards for attributes in stated-preference studies that include measurability (endpoints are quantifiable), interpretability (endpoints can be understood by a nonscientist), and comprehensiveness (all relevant endpoints are described).

Once the attributes have been defined, attribute levels must be specified. In some cases this is simple, such as the presence or absence of some attribute. In other cases the assignment of levels is more difficult, such as determining the appropriate levels and ranges used to specify forest species richness (Horne et al. 2005). This issue is also faced when specifying price or cost levels. Because the price/cost attribute provides control over the key factor that determines welfare measures, it is important that this attribute can be estimated precisely in the econometric model and also be reasonable in the policy context. Much as in contingent valuation, we would like low-price/cost alternatives to be frequently purchased and high-price alternatives to be rarely purchased.[4] Price levels should not be so high or low that they do not appear to be credible, but it may be informative for prices/costs to lie outside the range of existing market prices (such as travel costs) or be reasonable costs for the provision of public programs. Pilot studies play an important role in testing price or cost levels, as well as all other attributes and levels, to ensure that they have sufficient variation to identify the parameters and to ensure that welfare measures can be calculated.

These first two steps, which are critical to the successful implementation of CEs, are often not given the due consideration they require. Practitioners are encouraged to spend significant time and effort in scoping the problem, using focus groups and pretests, and making sure the choice context and scenario descriptions are carefully developed.

5.2.3 *Develop an Experimental Design*

Once attributes and levels have been determined, the researcher must determine the number of alternatives to present in each choice set (two, three, four, etc.), and the number of choice sets to present to the respondents (one, four, eight, 16, etc.). The number of alternatives could depend on the type of value being measured and/or on the context of the study. At a minimum, choice questions should contain a status quo alternative and an alternative indicating a change from the status quo. A status quo alternative is required in each choice set so that estimated utility functions represent changes from baseline conditions. Total value (or passive-use value) studies often employ only two alternatives because of the incentive compatibility of a two-alternative choice or referendum (Carson and Groves 2007). The number of

[4]A useful graphical tool for visualizing the role of price on choice in a multiattribute context is described by Sur et al. (2007).

alternatives in some studies depends on the number of alternatives that occur in the real world.

The number of choice questions to ask depends in part on the complexity of the task and is often a judgment the researcher must make based on focus groups, pilot tests, and expert judgment. In general, the number of choice sets included in the design depends on the number of degrees of freedom required to identify the model. The use of multiple choice sets can also have implications for incentive compatibility (Carson and Groves 2007, 2011).

Experimental design procedures are used to assign attribute levels to the alternatives that form the basis for choices and to construct the sets of choices that will be presented to respondents. Alternatives presented to the respondents must provide sufficient variation over the attribute levels to allow one to identify preference parameters associated with the attributes. In most cases, presenting all combinations of attributes and levels will be impossible. Thus, experimental design procedures are used to identify subsets of the possible combinations that best identify attribute preferences. Because of the importance of this topic to the success of any CE (Scarpa and Rose 2008), it is discussed in detail in Sect. 5.3.

5.2.4 Questionnaire Development

As with other stated-preference methods, CEs involve surveys, and various questionnaire formats can be used for collecting data (see Chap. 3), including:

- Mail-out, mail-back surveys.
- Telephone recruitment, mail-out, mail-back surveys.
- Telephone recruitment, mail-out, telephone surveys.
- Computer-assisted surveys at centralized facilities or in person.
- Intercept surveys that could be paper and pencil or computer-assisted.
- Internet-based surveys, including Internet panels.

The selection of the questionnaire format is usually based on pragmatic concerns, such as availability of a sample frame and budget limitations. In the case of CEs, Internet modes, particularly Internet panels, are becoming increasingly popular. Because CEs present respondents with complex sets of choice questions and randomization of the order of these questions is desirable, mail and telephone surveys can be more difficult to use relative to Internet or computer-based in-person surveys (e.g., using tablets to collect information from respondents). Also, in some cases information from early parts of a survey is used in the design of attributes and/or levels in the choice tasks, making computer-based Internet or in-person surveys more convenient. While Chap. 3 discusses survey mode comparisons and trade-offs in general, specific features of CEs, such as the incorporation of the experimental design into the survey, are facilitated by the use of Internet panels. Concerns about the social context induced by in-person surveys (social desirability bias) and the cost of in-person surveys result in less use of this mode.

Lindhjem and Navrud (2011) reviewed survey mode effects in stated-preference models and found relatively little difference between Internet modes and other modes. However, they raised concerns about the representativeness of Internet modes. Some Internet panels have very good properties in terms of representativeness because they are based on random samples of the population, while others are opt-in panels that raise questions about selection bias.

Various methods can be used to communicate information about the attributes of a valuation problem. In addition to verbal descriptions, maps, and photographs, other graphic displays should be considered. As in any survey-based research, pretesting of the questionnaire is absolutely necessary to assure that respondents clearly understand the information being communicated (see Chap. 3 for more detail on survey methods). Choice experiments are typically presented as matrices with alternatives as the columns and attributes as the rows, but there are various other forms of presentation that can be used. Often researchers will include graphics or other visual aids within the choice matrix to represent attributes and levels.

An issue that is critical in the design of stated-preference surveys, including CEs, is the inclusion of methods to address hypothetical bias (strategic behavior) and, in some cases, to assess scope effects. CEs are probably as prone to strategic behavior or hypothetical bias as are contingent valuation tasks. Carson and Groves (2007, 2011) outlined the theory associated with strategic behavior and emphasized the need to construct instruments that are "consequential." They also described the ways that CEs can differ from other stated-preference methods in terms of strategic behavior. For example, strategic behavior can arise from the examination of choice sequences to look for the "best deal" (Holmes and Boyle 2005; Day et al. 2012), as well as from the design of the CE (Vossler et al. 2012).

Three major approaches for dealing with hypothetical bias in stated-preference surveys have been used. The first is to include "cheap talk scripts" (Cummings and Taylor 1999; List 2001) that describe to respondents that hypothetical values or bids are often higher than they would be when there are real consequences. This approach is no longer being recommended because it may influence values by suggesting that respondents' values are often too high. Reminders of substitutes represent good practice, but statements that may influence values before the valuation question is asked are questionable (see Vossler, 2016, for details).

The second approach is to ask respondents how certain they are about their choice (Blumenschein et al. 2008). Uncertain preferences for a program (particularly regarding a nonstatus quo program in a passive-use context) could lead to status quo choices, and adjustments for uncertain preferences in CEs have been investigated (Ready et al. 2010). Finally, Vossler et al. (2012) outlined how choice experiments can generate strategic responses and showed that when respondents think that the program being presented is consequential (could actually be used in policy), responses are more likely to correspond to preferences elicited in an incentive-compatible fashion.

5.2.5 Data Collection

Data collection should be carried out using the best survey practices (e.g., Dillman 1978). Chapter 4 outlines a number of issues in data collection for contingent valuation studies that apply as well to the implementation of CEs. One unique feature arising in CEs is that multiple choice sets are presented to individuals with the intent that choice sets be considered independently and without comparing strategically across choice sets. This means that it is desirable to prevent respondents from reading ahead or going back and changing responses. It is also valuable to randomize the order of the presentation of the choice sets so that the first task, in a large enough sample, can be used to estimate values that are not affected by repeated choices. In a mail survey (paper and pencil), this is very difficult to accomplish because respondents can flip through the survey booklet. Computer-based surveys (Internet and in-person) can achieve this through the design of the survey implementation program. Computer-based methods also capture the amount of time spent on each question, which tells researchers if respondents are taking time to consider the choice set carefully.

5.2.6 Model Estimation

Once data have been collected, the next step is to estimate preference parameters using a random utility model. A growing number of econometric specifications have been used to analyze choice data. These models typically vary over how the error term is interpreted, particularly in the context of heterogeneity in preferences across respondents. Due to the complexity of these models and the variety of econometric specifications available, estimation is discussed in detail in Sects. 5.4 through 5.6.

5.2.7 Policy Analysis and Decision Support

Most CE applications are targeted to generating welfare measures (see Sect. 5.5), predictions of behavior, or both. These models are used to simulate outcomes that can be used in policy analysis or as components of decision support tools. CEs provide the opportunity to evaluate the welfare effects of multiple policy options involving combinations of attributes and levels. They also allow for calibration to actual policies or outcomes when these conditions become known. For example, choice experiments on park visitation have been calibrated using actual visitation information when measuring nonmarginal welfare impacts (Naidoo and Adamowicz 2005). As such, they can provide a richer set of policy information than most other valuation approaches.

Yellowstone National Park Case Study

PROBLEM

Debates over how national parks within the United States should be managed have been very divisive, with some groups arguing in favor of strict wilderness conservation while other groups have stressed the need for various in-park recreational activities. This debate has been especially intense in Yellowstone National Park, the oldest and arguably most well-known national park in the U.S. During the winter months, snowmobiles are the primary means of accessing the park. People who oppose the use of snowmobiles in the park claim that they are noisy and smelly, cause congestion and interfere with other recreational uses such as hiking and cross-country skiing, and threaten the park's wildlife. Proponents of snowmobile use in the park argue that the machines are safe, convenient, and often the only way to access the extraordinary winter landscape. Designing efficient and equitable winter use regulations for Yellowstone National Park has been a challenge for several decades.

APPROACH

A winter visitation survey was designed to obtain information relevant to a benefit-cost analysis of winter management alternatives being considered by the National Park Service. The sampling frame was based on users of the park during the winter of 2002-03. In addition to information on trip-related expenditures and winter recreation activities pursued, the survey implemented a choice experiment that was designed to solicit preferences for attributes of Yellowstone National Park under different winter management alternatives, including various snowmobile restrictions.

RESULTS

Study results indicated that some restrictions on snowmobile access in Yellowstone National Park could improve social welfare. Welfare losses to snowmobile riders could be offset by welfare gains to other park users although the net benefits depend on the number of riders and nonriders using the park as well as the specific regulations imposed. Further, heterogeneous preferences were found regarding the restriction for snowmobilers to be on a group tour—experienced snowmobilers did not like the restriction while novice snowmobilers did not mind group tours. The results of the survey have been used in several benefit-cost analyses to support rulemaking in Yellowstone National Park.

SOURCE

Mansfield, C., Phaneuf, D. J., Johnson, F. R., Yang, J.-C. & Beach, R. (2008). Preferences for public lands management under competing uses: The case of Yellowstone National Park. Land Economics, 84, 282-305.

5.3 Experimental Design

The basic problem addressed in the experimental design literature for CEs—given the selected attributes and their levels—is how to allocate attribute levels to alternatives and choice sets. Several approaches to experimental design for CEs have been proposed, and the best approach to use depends on which preference parameters need to be estimated and whether or not prior information on the parameters is available. The researcher also needs to think about the complexity of the design because the inclusion of many alternatives and choice sets can cause respondents to use decision-making shortcuts (heuristics) that might not reflect their true preferences.

An experimental design must contain sufficient independent variation among attribute levels within and across alternatives so that each preference parameter can be identified. For example, if the levels of an attribute are always identical across alternatives, it will not be possible to identify the effect of that attribute on responses. A good design is also statistically efficient, meaning it minimizes (maximizes) the standard errors (precision) of the preference parameter estimates.

Traditional experimental designs were constructed to support linear-in-parameters statistical models, and orthogonal designs were popularized because they eliminated correlation between attributes so that the independent influence of each variable on outcomes could be estimated. Researchers have come to realize that orthogonal designs might not, in most situations, provide optimal statistical efficiency when nonlinear-in-parameters models are used to analyze CE data. This section provides an overview of traditional orthogonal designs as well as statistically efficient designs that are being adopted by CE practitioners.

5.3.1 *Orthogonal Full Factorial Designs*

The most complete experimental design is a full factorial design, which combines every level of each attribute with every level of all other attributes (Hensher et al. 2005). The primary advantage of a full factorial design is that all main and interaction effects are statistically independent (orthogonal) and can be identified when estimating a model.[5] The major drawback of this design is that a very large number of alternatives are generated as the numbers of attributes and levels are increased. For example, suppose that a recreation agency is evaluating plans for developing

[5]A main effect is the direct effect of an attribute on a response variable (choice), and it reflects the difference between the average response to each attribute level and the average response across all attributes (Louviere et al. 2000). An interaction effect occurs if the response to the level of one attribute is influenced by the level of another attribute. Interaction effects are represented by parameter estimates for the interaction (cross product) of two or more variables and can account for more complex behavioral responses to combinations of attribute levels.

Table 5.2 Orthogonal codes illustrating the properties of full and fractional factorial designs

	Main effects			**2-way interactions**			**3-way interactions**
Attribute combination	Picnic shelters ($A1$)	Boat ramps ($A2$)	Camping fee ($A3$)	$A1 \times A2$	$A1 \times A3$	$A2 \times A3$	$A1 \times A2 \times A3$
First fraction							
1	−1	−1	+1	+1	−1	−1	+1
2	−1	+1	−1	−1	+1	−1	+1
3	+1	−1	−1	−1	−1	+1	+1
4	+1	+1	+1	+1	+1	+1	+1
Second fraction							
5	−1	−1	−1	+1	+1	+1	−1
6	−1	+1	+1	−1	−1	+1	−1
7	+1	−1	+1	−1	+1	−1	−1
8	+1	+1	−1	+1	−1	−1	−1

lakeside campgrounds and that agency managers are considering whether or not to install picnic shelters and boat ramps at each location as well as how much to charge for an overnight camping fee. If each of these three attributes takes two levels (install or not for facilities, \$10 or \$20 for camping fees), there are $2^3 = 8$ possible combinations of attribute levels in the full factorial design (the eight alternatives shown in Table 5.2).[6] If the number of levels associated with each attribute increases from two to three, the full factorial design increases to $3^3 = 27$ possible combinations of attribute levels.

The properties of an experimental design can be understood using a helpful coding scheme known as "orthogonal coding." Under this scheme, attribute levels are assigned values so that the sum of values in each column (representing main or interaction effects) equals zero (Hensher et al. 2005). For an attribute with two levels, for example, this is accomplished by assigning a value of 1 for the first level of the attribute and a value of −1 for the second level. As is illustrated in Table 5.2, the sum of values for each main (or interaction) effect is zero because each level appears equally often. A design with this property is said to be balanced, and it ensures that preference parameters are well-estimated across the range of each attribute.[7] Further, Table 5.2 shows that the inner product of any two column vectors equals zero. This is because each pair of levels appears equally often (in

[6]More generally, the number of possible combinations of attribute levels in a full factorial design is $\pi_{k=1}^{K} L_k$, where L_k is the number of attribute levels associated with attribute k.

[7]Attribute level balance leads to larger experimental designs when the number of attribute levels differs across attributes.

one-quarter of the attribute combinations) across all columns and indicates that all of the main effects are orthogonal.[8]

Orthogonal codes are also useful for identifying the design columns associated with interaction effects. In Table 5.2, let Attribute 1 ($A1$) represent picnic shelters, Attribute 2 ($A2$) represent boat ramps, and Attribute 3 ($A3$) represent camping fees. Then, the design of all interaction effects is found by multiplying together the appropriate orthogonal codes associated with each attribute. In the first alternative, for example, the two-way interaction between picnic shelters and boat ramps ($A1 \times A2$) is computed as ($-1 \times -1 = 1$). By computing the inner products of the columns of orthogonal codes, one can see that all interaction effects are statistically independent of each other and also independent of the main effects.

5.3.2 Orthogonal Fractional Factorial Designs

The number of attribute combinations needed to represent a full factorial design increases rapidly as the number of attributes and levels increases, and fractional factorial designs can be used to reduce the design size. The simplest method of generating a fractional factorial design is to select subsets of attribute combinations from the full factorial design using higher order interaction terms (Hensher et al. 2005). For example, in Table 5.2, two fractions of the full factorial are shown based on the three-way interaction ($A \times B \times C$) which takes the value +1 for the first half-fraction and −1 for the second half-fraction. Note that for each half-fraction (four alternatives), the design remains balanced and orthogonal for the main effects.

However, in reducing the design size, fractional factorial designs omit some or all information on interaction effects. If the omitted interactions are important in explaining responses, the preference parameter estimates may be biased due to confounding an omitted variable with a main effect. To see this (Table 5.2), note that the vector of two-way interactions $A1 \times A2$ [+ 1, −1, −1, +1] is identical to the vector of main effects for $A3$. Thus, $A1 \times A2$ is perfectly collinear with $A3$. If econometric analysis of data collected using this fractional factorial design showed that the parameter estimate on $A3$ was significantly different than zero, we could not be sure whether camping fees ($A3$) were significant, the interaction of picnic shelters and boat ramps was significant ($A1 \times A2$), or both. Thus the interpretation of the parameter estimates (mean and standard error) on the key economic variable (camping fee) would be ambiguous because a potentially important interaction between facility attributes was not identified.

[8]In general, orthogonality occurs when the joint occurrence of any two attribute levels, for different attributes, appear in attribute combinations with a frequency equal to the product of their individual frequencies. In Table 5.2 for example, each attribute level (−1 or 1) for each attribute appears in one-half of the attribute combinations. Therefore, the joint combination of any two attribute levels (say, −1 and −1) must occur in ½ × ½ = ¼ of the attribute combinations for the design to be orthogonal.

Due to the possibility that a fractional factorial design can induce biased parameter estimates on the main effects and can fail to identify meaningful interactions, it is essential to identify which attribute interactions might be important during the design stage of survey development. For example, a CE that is investigating alternative transportation routes might anticipate that some combination of attributes, such as low travel cost and low road congestion, are strongly preferred. These potentially important interactions could be evaluated by asking focus group participants if some combinations of attribute levels are particularly desirable or undesirable. If so, a main effects plus selected interactions plan should be used. In general, this is accomplished using orthogonal codes to examine the correlations between the main and interaction effects and assigning attributes to design columns that are orthogonal to the specified interaction effects (Hensher et al. 2005).

5.3.3 Generating Choice Sets for Orthogonal Designs

The key issues to consider in creating an experimental design for a CE are how to place alternatives into choice sets and how many choice sets are needed. Several choice sets are typically included in a CE, and the number of choice sets depends on the number of degrees of freedom (the number of parameters plus one) required to identify the parameters of the specified model. In turn, the number of degrees of freedom depends on whether the alternatives are described using a label to differentiate the alternatives (such as transportation modes or recreational locations) or whether they are unlabeled (generic). Labeled alternatives are used when the researcher wants to estimate a utility function for each alternative. The ability to identify the independent effect of each attribute in each alternative requires that attributes are orthogonal within and between alternatives. Unlabeled alternatives are used when the researcher is only interested in estimating a single utility function. Because labeled designs require more parameters to be estimated, more degrees of freedom are required, resulting in a larger design.

Note that labeled designs permit the analyst to estimate a constant term specific to each alternative, known as an alternative specific constant. If respondents derive utility from unobserved attributes associated with the labels of alternatives, the alternative specific constants provide a means for measuring that component of utility that is independent of the experimentally designed attributes. It is also common to test for status quo bias in unlabeled designs by including an alternative specific constant for the status quo alternative. If the alternative specific constant is statistically significant, it suggests that respondents have a preference for (or against) the status quo option independent of the designed attributes.

The number of degrees of freedom also depends on whether parameters are estimated for the levels of an attribute (referred to as nonlinear effects) or whether a single parameter is estimated for an attribute (referred to as a linear effect).[9] For categorical variables, it is necessary to estimate nonlinear effects, and the number of nonlinear effects that can be estimated equals the number of levels (such as low, medium, or high) minus one (hence two nonlinear effects). For continuous variables such as price, a single parameter is usually estimated.

The number of attribute combinations required to estimate main effects must equal or exceed the number of degrees of freedom, which can be simply computed (Hensher et al. 2005). For unlabeled alternatives in which the analyst wants to estimate nonlinear effects, the minimum degrees of freedom required is $(L - 1) \times A + 1$, where L is the number of attribute levels and A is the number of attributes. If only one parameter is estimated for each attribute, the number of degrees of freedom is reduced to $A + 1$. For labeled alternatives, the comparable formulas are $(L - 1) \times NA + 1$ and $NA + 1$, where N is the number of alternatives.[10]

Experimental designs for unlabeled alternatives can be created starting with an orthogonal plan of attribute combinations. Each attribute combination (such as combinations 1 through 4 in Table 5.2) provides a design for the first alternative in one of four choice sets. Then a second alternative could be created by randomly pairing nonidentical attribute combinations (Table 5.3). This design is suitable for estimating the three desired parameters (one parameter for each attribute) because there are four degrees of freedom (the number of parameters plus one) and four sets of attribute combinations. However, note that the design codes for specific attributes are identical in some attribute combinations (such as picnic shelters in the first combination). Because choice models are based on attribute differences, the lack of a contrast for attribute levels within choice sets reduces the statistical efficiency of the design. Also note that this design would be unsuitable for estimating the parameters of a labeled experiment because there is an insufficient number of attribute combinations relative to the degrees of freedom ($2 \times 3 + 1 = 7$) required to identify each parameter. Further, as can be seen by computing the inner product of the camping fee attribute levels in Alternatives A and B (Table 5.3), these columns are not orthogonal. In fact, they are perfectly (negatively) correlated. In general, randomizing attribute combinations will be inadequate for estimating independent utility functions for labeled alternatives because the attributes will not be orthogonal across alternatives, and main effects will be correlated (Hensher et al. 2005; Street et al. 2005).

---

[9]Nonlinear effects in this context should not be confused with functional forms of the variables, such as quadratic or logarithmic transformations. If the researcher is interested in whether continuous variables (such as price) are better described by nonlinear functional forms, nonlinear effects could be estimated and used to evaluate the functional form.

[10]If the number of attribute levels differs across attributes, then the formulas for computing the number of degrees of freedom required to estimate nonlinear effects must be adjusted. In particular, the value of $(L - 1) \times A$ must be computed for each set of attributes with a unique number of levels. Then these values must be summed before adding 1.

Table 5.3 Orthogonal codes illustrating properties of choice sets for random attribute combinations

	Alternative *A*				**Alternative *B***		
Attribute combination	Picnic shelters	Boat ramps	Camping fee	Random combination	Picnic shelters	Boat ramps	Camping fee
1	−1	−1	+1	2	−1	+1	−1
2	−1	+1	−1	4	+1	+1	+1
3	+1	−1	−1	1	−1	−1	+1
4	+1	+1	+1	3	+1	−1	−1

An alternative approach is to use a design based on a collective full factorial in which the design of choice sets occurs simultaneously with the design of alternatives, and the attribute levels are orthogonal within and across alternatives. A collective full factorial design is referred to as an L^{NA} design (Louviere et al. 2000). Using the campground example with $A = 3$ and $L = 2$, if one decides to include two alternatives plus the status quo alternative in each choice set, then the collective full factorial design includes $2^{(2 \times 3)} = 64$ rows, and each row contains a design for both of the new campground alternatives.[11]

Given a collective full factorial, a main effects fractional factorial design can be selected as the smallest orthogonal plan, and it depends on the number of degrees of freedom required to estimate all of the main effects. The smallest orthogonal design available is usually larger than the number of degrees of freedom required to estimate the parameters, and finding the smallest orthogonal design for differing levels of attributes and levels is mathematically challenging (Louviere et al. 2000). Fortunately, orthogonal main effects plans are available in published literature, on the Internet, and by using software programs.

In the campground example, the number of degrees of freedom required to estimate a main effects only collective fractional factorial design is $(2 - 1) \times (2 \; 3) + 1 = 7$. An example of a labeled campground design is shown in Table 5.4, and as can be seen, the smallest collective fractional factorial design includes eight choice sets. Each attribute (design column) is orthogonal to all other attributes, both within and across alternatives. Note that the design codes for specific attributes (such as picnic shelters) are identical in some sets (such as in Choice sets 1, 2, 5, and 6). Also note that some attribute combinations (Combinations 1 and 6) have identical orthogonal codes for each labeled alternative. In general, this design would not provide any information for those attribute combinations if the attribute levels associated with each code were identical. However, it is common practice when using labeled designs to assign different values for the attribute levels associated with each alternative. For example, the camping fees associated with Label A might be \$10 (coded −1) versus \$20 (coded +1), and the camping fees associated

[11] As the attribute levels of the status quo alternative are held constant across choice sets, the status quo alternative is not included in N (the number of alternatives).

Table 5.4 Orthogonal codes illustrating properties of choice sets for a labeled, collective fractional factorial design

	Label A			Label B		
Attribute combination	Picnic shelters	Boat ramps	Camping fee	Picnic shelters	Boat ramps	Camping fees
1	+1	+1	+1	+1	+1	+1
2	+1	+1	−1	+1	−1	−1
3	+1	−1	+1	−1	−1	−1
4	+1	−1	−1	−1	+1	+1
5	−1	+1	+1	−1	−1	+1
6	−1	+1	−1	−1	+1	−1
7	−1	−1	+1	+1	+1	−1
8	−1	−1	−1	+1	−1	+1

Table 5.5 A labeled main effects campground choice set taken from a $2^{(2 \times 4)}$ collective fractional factorial

	Campground A	**Campground B**	**Status quo alternative**
Picnic shelters	Yes	Yes	Stay at home: I would not choose either campground ☐
Boat ramps	Yes	Yes	
Camping fee	$20	$25	
I would choose: (please check one box)	☐	☐	

with Label B might be $15 (coded −1) versus $25 (coded +1). Thus, these alternatives would be contrasted on the camping fee attribute.

Once the design of the alternatives to be included in choice sets has been accomplished, the analyst needs to translate the orthogonal (or other design) codes into a format that makes sense to respondents and to add the status quo alternative. An example of a labeled choice set (based on the first attribute combination in Table 5.4) as it might appear in a questionnaire is shown in Table 5.5.

If the alternatives shown in Table 5.5 were unlabeled (generic), the first alternative would dominate the second alternative because both alternatives offer identical facilities and the first alternative is less expensive. Dominated alternatives are undesirable because they do not provide any useful information and therefore reduce the statistical efficiency of the design. However, for labeled alternatives, if unobserved attributes associated with the second alternative (Campground B) contributed substantially to respondent utility, that alternative might be chosen. Therefore, dominated alternatives are more difficult to detect for labeled alternatives.

Although it is not generally known how many choice sets respondents can evaluate, it is common to present four to six choice sets per respondent; sometimes 16 or more sets are included. If it is thought that the number of designed choice sets will place too great a cognitive burden on respondents, the burden can be reduced by assigning respondents to blocks, or subsets, of the fractional factorial design. One method used for blocking is to list the choice sets in random order and then subdivide the list to obtain blocks of reasonable size. A second method of blocking is to consider blocks as a separate attribute in the experimental design, letting the number of levels represent the desired number of blocks. This second method is preferred because including blocks as attributes in an orthogonal design assures that every level of all attributes will be present in every block (Adamowicz et al. 1998).

5.3.4 Statistical Efficiency for CEs

Traditional orthogonal designs were developed for linear-in-parameters statistical models and meet two criteria for good designs: (1) they remove multicollinearity among attributes so that the independent influence of each attribute can be estimated, and (2) they minimize the variance of the parameter estimates so that *t*-ratios (based on the square roots of the variances) are maximized. These design criteria are met when the elements of the variance-covariance matrix for the linear model are minimized (Rose and Bliemer 2009).[12]

For linear-in-parameters statistical models, the information used to estimate the variance-covariance matrix depends on the design matrix (data on the explanatory variables of the model) and a constant scaling factor. In contrast, the variance-covariance matrix for the nonlinear models used to analyze CE data contains information on the design matrix and the preference parameters (McFadden 1974).[13] A statistically efficient CE design has the best variance-covariance matrix. Various definitions of "best" have been proposed, and they depend on the assumptions made about the preference parameters as well as the method chosen to summarize information in the variance-covariance matrix.

A commonly used summary statistic for the information contained in the variance-covariance matrix is the determinant as it uses information on the main diagonal (variances) and the off-diagonals (covariances). The determinant of a variance-covariance matrix, scaled by the number of parameters to be estimated in the model, is known as the *D*-error. Designs that minimize the *D*-error are considered to be *D*-efficient.[14]

[12]The variance-covariance matrix is the inverse of the Fisher information matrix and is based on the second derivative of the log-likelihood function.

[13]In particular, McFadden (1974) showed that $\mathrm{VC} = \left[\sum_{n=1}^{N} \sum_{j=1}^{J_n} x'_{jn} P_{jn}(Z, \beta) x_{jn}\right]^{-1}$, where P_{jn} is the probability that an individual will choose Alternative j in Choice set n, which is a function of the attribute design matrix (Z) and a vector of preference parameters (β). Also, $x_{jn} = z_{jn} - \sum_{i=1}^{J_n} z_{in} P_{in}$, where z_{jn} is a row vector describing the attributes of Alternative j in Choice set n.

[14]Other criteria for design efficiency have been proposed in the literature. For example, the *A*-error minimizes the trace of the variance-covariance matrix, which is computed as the sum of the elements on the main diagonal.

5.3.4.1 Optimal Orthogonal Designs

One approach for finding D-efficient designs for CEs is to assume that all alternatives contained in choice sets are equally attractive or, equivalently, that all preference parameters equal zero.[15] These designs are referred to as optimal orthogonal designs.

An optimal orthogonal CE design is initialized with an orthogonal design for the first alternative in a choice set; it then makes systematic attribute level changes in the design to generate other alternatives (Street et al. 2005; Street and Burgess 2007).[16] Optimality for these designs is defined by two criteria: (1) the attributes within alternatives are orthogonal, and (2) the number of times an attribute takes the same level across alternatives in a choice set is minimized (known as the minimal overlap property). Under the second criterion, survey respondents must make trade-offs on all attributes in a choice set that, presumably, provides more information about preferences and avoids dominated alternatives, which can arise in traditional orthogonal designs.[17]

An optimal orthogonal design for our campground example is illustrated in Table 5.6. Beginning with an orthogonal design for Alternative A (as in Table 5.3), Alternative B was created by multiplying the levels in Alternative A by -1.[18] This fold-over procedure maintains the orthogonality of the design, which is also balanced, while providing a contrast for each level of each attribute. This procedure obviously would not work for labeled designs because the attributes in each alternative are perfectly (negatively) correlated.

In general, a D-efficient optimal orthogonal design is constructed by minimizing the following expression:

$$D_0\text{-error} = \det(\mathrm{VC}(Z, 0))^{1/k}, \tag{5.1}$$

where Z represents the attributes in the experimental design, 0 indicates that $\beta = 0$ for all model parameters, and k is the number of parameters used in the scaling factor. The efficiency of the fold-over design (Table 5.6) relative to the design obtained using random attribute combinations (Table 5.3) can be demonstrated by computing Eq. (5.1) for each design. In particular, the authors find that the D_0-error

[15]Huber and Zwerina (1996) showed that, under the assumption that $\beta = 0$, the variance-covariance matrix simplifies to $\left[\sum_{n=1}^{N}\frac{1}{J_n}\sum_{j=1}^{J_n}x'_{jn}x_{jn}\right]^{-1}$, where $x_{jn} = z_{jn} - \frac{1}{J_n}\sum_{i=1}^{J_n} z_{jn}$.

[16]This procedure, referred to as a "shifted design," was initially proposed by Bunch et al. (1996). In general, these designs use modulo arithmetic to shift the original design columns so they take on different levels from the initial orthogonal design.

[17]This approach implicitly assumes that the cognitive burden imposed by making difficult trade-offs does not influence the error variance and, therefore, does not bias parameter estimates.

[18]To use modulo arithmetic in constructing Table 5.6, begin by recoding each of the −1 values as 0. Then, add 1 to each value in Alternative A except for attributes at the highest level (1), which are assigned the lowest value (0).

Table 5.6 Orthogonal codes illustrating optimal orthogonal pairs using a fold-over design

	Alternative *A*			Alternative *B*		
Attribute combination	**Picnic shelters**	**Boat ramps**	**Camping fee**	**Picnic shelters**	**Boat ramps**	**Camping fee**
1	−1	−1	+1	+1	+1	−1
2	−1	+1	−1	+1	−1	+1
3	+1	−1	−1	−1	+1	+1
4	+1	+1	+1	−1	−1	−1

equals 0.79 for the random attribute combinations, and it equals 0.5 for the fold-over design, indicating the superiority of the latter design.[19]

5.3.4.2 Nonzero Priors Designs

A second approach to the efficient design of CEs using the variance-covariance matrix is based on the idea that information about the vector of preference parameters might be available from pretests or pilot studies and that this information should be incorporated in the design (Huber and Zwerina 1996; Kanninen 2002; Carlsson and Martinsson 2003; Hensher et al. 2005; Scarpa and Rose 2008; Rose and Bliemer 2009).[20] This approach, which we call a nonzero priors design, seeks to minimize the following expression:

$$D_p\text{-error} = \det(\mathrm{VC}(Z, \beta))^{1/k}, \tag{5.2}$$

where p stands for the point estimates of the (nonzero) β's. The constraints imposed on the optimal orthogonal model (orthogonality, attribute level balance, and minimal overlap) are relaxed in minimizing the D_p-error. However, if reasonable nonzero priors are available, relaxing these constraints can result in efficient designs that greatly reduce the number of respondents needed to achieve a given level of significance for the parameter estimates (Huber and Zwerina 1996). Note that designs that minimize the D_p-error do not generally minimize the D_0-error and vice versa.

If the nonzero priors used in Eq. (5.2) are incorrect, however, the selected design will not be the most efficient. One method for evaluating this potential shortcoming is to test the sensitivity of a D-efficient design to alternative parameter values, which can provide the analyst some degree of confidence about the robustness of a design (Rose and Bliemer 2009). Another approach that can incorporate the analyst's uncertainty about parameter values is to specify a distribution of plausible values

[19]Although the covariances equal zero in both designs, the efficiency of the fold-over design is gained by the minimal overlap property.

[20]One approach to developing nonzero priors is to use an orthogonal design in a pilot study to estimate the β vector, which is then used to minimize the D_p-error (Bliemer and Rose 2011).

that reflects subjective beliefs about the probabilities that specific parameter values occur (Sándor and Wedel 2001; Kessels et al. 2008). This Bayesian approach to experimental design proceeds by evaluating the efficiency of a design over many draws from the prior parameter distributions $f(\tilde{\beta})$. The design that minimizes the expected value of the determinant shown in Eq. (5.3) is a D-efficient Bayesian design:

$$D_b\text{-error} = \int_{\tilde{\beta}} \det\left(\mathrm{VC}(Z,\tilde{\beta})\right) f(\tilde{\beta})\mathrm{d}\beta \tag{5.3}$$

The distribution of $f(\tilde{\beta})$ is typically specified as normal or uniform.

Note that a nonzero priors design that is efficient for estimating one model (such as a multinomial logit model) is not necessarily efficient for estimating other models (such as random parameter logit or latent class models), and efforts are being made to identify designs that are robust to alternative model types (Ferrini and Scarpa 2007; Rose and Bliemer 2009). Also of interest is the construction of efficient experimental designs for the estimation of willingness to pay (WTP) measures, which are computed as the ratio of two parameters (Scarpa and Rose 2008). Because efficient designs can increase the cognitive burden faced by respondents by requesting them to make difficult choices, understanding the trade-offs between statistical efficiency and response efficiency is an emerging area of concern (Louviere et al. 2008; Johnson et al. 2013).

5.3.5 Selecting a Design

Given a suite of alternative design options, which design should a researcher choose? Although this will depend on considerations specific to each study, the authors recommend the following general guidelines. First, use a design that is statistically efficient in the context of the nonlinear-in-parameters models used to analyze random utility models. If reasonable information is available on preference parameters from sources such as pretests or pilot studies, the authors recommend using a nonzero priors design. In general, these designs reduce the number of respondents needed to achieve a specific precision (standard error) for the parameters specified in the utility function(s) and can therefore help reduce the cost of survey implementation. In cases where no prior information is available or where parameter estimates from other CE studies do not provide a good match, an optimal orthogonal design should be considered. This recommendation is based on evidence that optimal orthogonal designs can produce good results where prior information on parameter values is of poor quality or when the model specification chosen by the analyst is inconsistent with the underlying data generating process (Ferrini and Scarpa 2007). The construction of statistically efficient designs is greatly facilitated by the availability of software programs (such as SAS and Ngene).

5.3.6 Attribute Coding Schemes

While orthogonal codes are used to study the properties of an experimental design, other codes are used at the data analysis stage. Recall that attributes can be coded to estimate either linear or nonlinear effects (Sect. 5.3.3). For continuous variables, such as the cost of an alternative, it is common to estimate a linear effect by using the level of the quantitative variable as the code. However, it is also possible to estimate the nonlinear effects of a continuous variable using a nonlinear effects coding method that is also used for coding qualitative attributes. Two coding methods are available for estimating nonlinear effects. First, dummy variables, using $0 - 1$ codes, can be defined for $L - 1$ levels of an attribute. However, no information is recovered about preferences for the omitted level because all of the omitted levels are confounded. This limitation can be overcome using a second method known as "effects codes," which creates a unique base level for each attribute (Louviere et al. 2000).

Begin by creating an effects-coded variable EC_1, for the first attribute using the following steps:

- If the profile contains the first level of the attribute, set $EC_1 = 1$.
- If the profile contains the Lth level of the attribute, set $EC_1 = -1$.
- If neither Step 1 nor Step 2 apply, set $EC_1 = 0$.

If an attribute has two levels, one only needs to create one effects-coded variable using the preceding three steps for that attribute. However, if an attribute has three levels, one continues the coding process by creating a second effects-coded variable, EC_2, for that attribute using three additional steps:

- If the profile contains the second level of the attribute, set $EC_2 = 1$.
- If the profile contains the Lth level of the attribute, set $EC_2 = -1$.
- If neither Step 4 nor Step 5 apply, set $EC_2 = 0$.

If an attribute has more than three levels, one continues creating effects codes in this manner until $(L - 1)$ effects codes are created for each L-level attribute. Using this coding scheme, the parameter value for the omitted attribute level can be simply computed. For example, the value of the parameter for the Lth level of an attribute is the sum $b_1(-1) + b_2(-1) + \cdots + b_{L-1}(-1)$, where b_n is the parameter estimate on the nth level $(n \neq L)$ of an effects-coded variable.

For labeled experiments, as well as for the status quo alternative in a generic experiment, it is important to include a code for the alternative specific constant. Alternative specific constants are coded using dummy variables and, if there are N alternatives in the choice set, then $(N - 1)$ alternative specific constants can be included in the econometric specification. Because the status quo alternative will typically set the attributes at their current level (unless the status quo is an alternative such as "stay at home"), the status quo levels are coded with the same codes used for the other alternatives.

5.4 The Random Utility Model

The analysis of responses to a choice experiment is based on an extension of the random utility maximization (RUM) model that underlies discrete choice contingent valuation responses (Chap. 4) and recreation site choices between competing alternatives (Chap. 6). The CE format focuses the respondent's attention on the trade-offs between attributes that are implicit in making a choice. As shown below, model estimates are based on utility differences across the alternatives contained in choice sets.

The RUM model is based on the assumption that individuals know their utility with certainty, but analysts are unable to perfectly observe respondent utility so the unobservable elements are part of the random error. This assumption is formalized in a model where utility is the sum of systematic (v) and random (ε) components for individual k:

$$V_{ik} = v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik}, \tag{5.4}$$

where V_{ik} is the true but unobservable indirect utility associate with Alternative i, Z_i is a vector of attributes associated with Alternative i, p_i is the cost of Alternative i, y_k is income, and ε_{ik} is a random error term with zero mean.

For simplicity, let's consider an individual who is faced with a choice between mutually exclusive alternatives, where each alternative is described with a vector of attributes, Z_i. We assume that this individual maximizes their utility when making a choice. Therefore the individual will choose Alternative i if and only if

$$v_{ik}(Z_i, y_k - p_i) > v_{jk}\left(Z_j, y_k - p_j\right); \quad \forall j \in C, \tag{5.5}$$

where C contains all of the alternatives in the choice set. However, from an analyst's perspective, unobserved factors that influence choice enter the error term and, thus, individual k will choose Alternative i if and only if

$$v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik} > v_{jk}\left(Z_j, y_k - p_j\right) + \varepsilon_{jk}; \quad \forall j \in C. \tag{5.6}$$

The stochastic term in the random utility function allows probabilistic statements to be made about choice behavior. The probability that a consumer will choose Alternative i from a choice set containing competing alternatives can be expressed as

$$P_{ik} = P\left[v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik} > v_{jk}\left(Z_j, y_k - p_j\right) + \varepsilon_{jk}; \quad \forall j \in C\right]. \tag{5.7}$$

Equation (5.8) is very general, and assumptions need to be made about the specification of the utility function and the probability distribution of the error terms in order to estimate a model.

5.4.1 Specification of the Utility Function

A common assumption is that utility is a linear function of the attributes included in the experimental design so that the utility of choosing Alternative i is

$$v_{ik} = \beta Z_i + \lambda(y_k - p_i) + \varepsilon_{ik}, \tag{5.8}$$

where β is the vector of preference parameters for nonmonetary attributes and λ is the marginal utility of money. When choosing a specification, there is a trade-off between the benefits of assuming a less restrictive formulation (e.g., including interaction terms) and the complications that arise from doing so. Furthermore, the specifications that can actually be identified depend on the experimental design (see Sect. 5.3).

Consider an experiment with three attributes, including a monetary attribute. A utility function that is a linear function of the attributes would then be written as

$$v_{ik} = \beta_1 z_{i1} + \beta_2 z_{i2} + \lambda(y_k - p_i) + \varepsilon_{ik}. \tag{5.9}$$

However, if the experiment allows for an interaction term between the two nonmonetary attributes, the utility function could be specified as

$$v_{ik} = \beta_1 z_{i1} + \beta_2 z_{i2} + \beta_3 z_{i1} z_{i2} + \lambda(y_k - p_i) + \varepsilon_{ik}. \tag{5.10}$$

Note that this function remains linear in parameters, but it is not a linear function of the attributes.

One important property of discrete choice models is that only the differences in utility between alternatives affect the choice probabilities—not the absolute levels of utility. This can be shown by rearranging the terms in Eq. (5.7):

$$P_{ik} = P\left[\varepsilon_{ik} - \varepsilon_{jk} > v_{jk}\left(Z_j, y_k - p_j\right) - v_{ik}(Z_i, y_k - p_i); \quad \forall j \in C\right]. \tag{5.11}$$

Here one sees that choices are made based on utility differences across alternatives. Thus, any variable that remains the same across alternatives, such as respondent-specific characteristics like income, drops out of the model. Although Eq. (5.11) indicates that there must be a difference between attribute levels for competing alternatives in order to estimate the preference parameters for the attributes, the levels of some attributes could be equal in one or several of the choice sets.[21]

The property that there must be a difference between alternatives also has implications for the possibility of including alternative specific constants. Because alternative specific constants capture the average effect on utility of factors that are

[21] As discussed in Sect. 5.3, when attribute levels are the same across alternatives within a choice set, they do not elicit respondent trade-offs and therefore are uninformative.

not explicitly included as attributes, only differences in alternative specific constants matter. As was described in Sect. 5.3.6, a standard way of accounting for this is to normalize one of the constants to zero so that the other constants are interpreted as relative to the normalized constant.

The fact that only utility differences matter also has implications for how socioeconomic characteristics can enter a RUM model. Socioeconomic characteristics are used to capture observed taste variation, and one way of including them in the model is as multiplicative interactions with the alternative specific constants. Otherwise, these characteristics could be interacted with the attributes of the alternatives.

5.4.2 The Multinomial Logit Model

The next step is to make an assumption regarding the distribution of the error term. Alternative probabilistic choice models can be derived depending on the specific assumptions that are made about the distribution of the random error term in Eq. (5.11). The standard assumption in using a RUM model has been that errors are independently and identically distributed following a Type 1 extreme value (Gumbel) distribution. The difference between two Gumbel distributions results in a logistic distribution, yielding a conditional or multinomial logit model (McFadden 1974). This model relies on restrictive assumptions, and its popularity rests to a large extent on its simplicity of estimation. The multinomial logit model is introduced first and its limitations are discussed before introducing less restrictive models.

For simplicity, suppose that the choice experiment to be analyzed consists of one choice set containing N alternatives ($i = 1, \ldots, N$). If errors are distributed as Type 1 extreme value, the multinomial logit model applies, and the probability of respondent k choosing Alternative i is

$$P_{ik} = \frac{\exp(\mu v_{ik})}{\sum_{j=1}^{N} \exp\left(\mu v_{jk}\right)}, \tag{5.12}$$

where μ is the scale parameter that reflects the variance of the unobserved part of utility (Ben-Akiva and Lerman 1985). In basic models, the scale parameter is typically set equal to one, although other formulations will be discussed below.

There are two important properties of the multinomial logit model: (1) the alternatives are treated as independent, and (2) the modeling of taste variation among respondents is limited. The first problem arises because of the independently and identically distributed assumption about the error terms and results in the famous independence of irrelevant alternatives property. This property states that the ratio of choice probabilities between two alternatives in a choice set is unaffected by other alternatives in the choice set. This can be seen in the expression for the ratio of choice probabilities for the multinomial logit model:

$$\frac{P_{ik}}{P_{nk}} = \frac{\exp(\mu v_{ik}) / \sum_{j=1}^{N} \exp(\mu v_{jk})}{\exp(\mu v_{nk}) / \sum_{j=1}^{N} \exp(\mu v_{jk})} = \frac{\exp(\mu v_{ik})}{\exp(\mu v_{nk})}. \tag{5.13}$$

This expression only depends on the attributes and the levels of the attributes for the two alternatives and is assumed to be independent of other alternatives in the choice set. This is a strong assumption that might not always be satisfied.

Fortunately, the assumption about independence of irrelevant alternatives can easily be tested. If independence of irrelevant alternatives is satisfied, then the ratio of choice probabilities should not be affected by whether another alternative is in the choice set or not. One way of testing independence of irrelevant alternatives is to remove one alternative and re-estimate the model and compare the choice probabilities. This type of test was developed by Hausmann and McFadden (1984) and is relatively simple to conduct (see Greene, 2002, for details). If the test indicates that the assumption of independence of irrelevant alternatives is violated, an alternative model should be considered. One type of model that relaxes the homoscedasticity assumption of the multinomial logit model is the nested multinomial logit model (Greene 2002). In this model, the alternatives are placed in subgroups, and the error variance is allowed to differ between the subgroups but is assumed to be the same within each group. Another alternative specification is to assume that error terms are independently, but nonidentically, distributed Type I extreme value, with a scale parameter (Bhat 1995). This would allow for different cross elasticities among all pairs of alternatives. Furthermore, one could also model heterogeneity in the covariance among nested alternatives (Bhat 1997).

The second limiting property of the multinomial logit model is how the model handles unobserved heterogeneity. As we will see, observed heterogeneity can be incorporated into the systematic part of the model by allowing for interaction between socio-economic characteristics and attributes of the alternatives or constant terms. However, the assumption about independently and identically distributed error terms is severely limiting with respect to unobserved heterogeneity.

5.5 Welfare Measures

The goal of most CEs is to estimate economic welfare for use in policy analysis. Because CEs provide quantitative measures of tradeoffs between attributes (including price), they can be used to estimate how much money respondents would be willing to pay for a change in attribute levels while remaining as well off after the change as they were before the change, which provides estimates of compensating variation (Chap. 2). The fact that CEs provide estimates of the indirect utility function allows one to calculate willingness to pay for gains or losses for any combination of change in attribute levels.

Section 5.3.3 described methods for generating choice sets for "state-of-the-world" (generic, unlabeled) experiments and alternative specific (labeled) designs, and it is important to understand that estimates of WTP are computed differently for these two designs. This is because in a state-of-the-world experiment, only one alternative can be realized at the end of the day. In an alternative specific experiment, several of the alternatives may exist at the same time. This means, in turn, that an additional problem arises because when one wants to make welfare evaluations, an assumption needs to be made about which alternative a respondent would choose. For example, changes in a policy attribute (such as adding a boat ramp at Campground A) might cause some respondents to choose that alternative instead of a different alternative, while some others already chose that alternative before the change and, finally, others will still not choose that alternative.

Another concept that is somewhat related to the difference between a state-of-the-world experiment and an alternative specific experiment is the difference between generic labels (such as Alternative *A*, Alternative *B*, and so forth) and explicit labels (such as "car," "bus"). The generic labels do not convey any information about the alternatives, so WTP is simply a function of preferences over the levels of the policy-related attributes. In contrast, explicit labels convey information about fixed attributes that are not associated with the experimentally designed attributes, and WTP is therefore a function of preferences regarding both the policy-related attributes and the fixed attributes associated with the label.

5.5.1 Willingness to Pay: State-of-the-World Experiments

In the case of a state-of-the-world CE, one can think that there is only a single alternative to consider that can exhibit various attribute conditions or "states of the world."

Assume a simple linear utility function for Alternative *i*, where the alternative simply represents a certain state of the world, and respondent *k*:

$$v_{ik} = \beta Z_i + \lambda(y_k - p_i) + \varepsilon_{ik}. \tag{5.14}$$

Suppose one wishes to estimate the respondent's WTP for a change in the attribute vector from initial conditions (Z_0) to altered conditions (Z_1). To estimate the compensating variation for a new state of the world versus a base case, one does not have to consider the probability of choosing different alternatives. Therefore, the compensating variation (CV) associated with this change is

$$\mathrm{CV} = \frac{1}{\lambda}\left\{V^1 - V^0\right\}, \tag{5.15}$$

where V_1 and V_0 are expressions of utility for the new and base case states of the world. For example, suppose one conducts a choice experiment with three attributes, including the cost attribute, and the following utility function is estimated:

$$v_{ik} = \beta_1 z_{i1} + \beta_2 z_{i2} + \lambda(y_k - p_i) + \varepsilon_{ik}. \tag{5.16}$$

Based on the estimated model, one wishes to calculate the WTP for changes in the two nonmonetary attributes relative to the base: Δz_{i1} and Δz_{i2}. WTP would then be

$$\mathrm{WTP} = -\frac{\beta_1 \Delta z_{i1} + \beta_2 \Delta z_{i2}}{\lambda}. \tag{5.17}$$

This expression shows the maximum amount of money an individual is willing to pay in order to obtain the improvement in the two attributes.

So far we have discussed WTP for a discrete change in multiple attributes. However, what is often reported from generic choice experiments is the marginal WTP. Using a simple linear utility function (Eq. 5.14), the marginal rate of substitution between any of the attributes and money is simply the ratio of the coefficient of the attribute and the marginal utility of money:

$$\mathrm{MRS} = -\frac{\partial v_{ik}/\partial z_i}{\partial v_{ik}/\partial y_k} = -\frac{\beta_i}{\lambda} = \mathrm{MWTP}. \tag{5.18}$$

Marginal WTP (also known as the implicit price) shows how much money an individual is willing to sacrifice for a marginal change in the attribute. Note that because the expression is a ratio of the coefficients, the scale parameter cancels from the expression.

However, in many instances the attributes are not continuous. For example, the attribute could be a dummy variable indicating if the attribute is present or not. In that case, the ratio of the attribute coefficient and the marginal utility of money is strictly not a marginal WTP because one cannot talk about a marginal change of the discrete attribute. The interpretation of this WTP measure is instead the amount of money a respondent is willing to pay for a change in the attribute from, say, not available to available.

Two extensions of the estimation of marginal WTP can now be considered. The first concerns nonlinear utility functions for which marginal WTP would have to be evaluated at a certain level of the attribute. Suppose one includes an interaction term in an experiment with two attributes so that the utility function is

$$v_{ik} = \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2 + \lambda(y_k - p_i) + \varepsilon_{ik}. \tag{5.19}$$

The marginal WTP for attribute z_1 is then

$$\mathrm{MWTP} = -\frac{\partial v_{ik}/\partial z_1}{\partial v_{ik}/\partial y_k} = -\frac{\beta_1 + \beta_3 z_2}{\lambda}. \tag{5.20}$$

Therefore, the marginal WTP for Attribute 1 depends on the level of the other attribute, and one would have to decide at what values to calculate the WTP.

The second extension is to allow for observed heterogeneity in WTP. This can be done by interacting attributes of the choice experiment with a set of socioeconomic characteristics (see Sect. 5.7.2). This way one would obtain the marginal WTP for different groups of people with a certain set of socioeconomic characteristics. Note that in many choice experiments, the socio-economic characteristics are interacting with the alternative specific constants. In that case they will not affect the marginal willingness to pay.

5.5.2 *Sources of Variation in Marginal WTP and Comparison of Models*

The expressions derived for marginal willingness to pay are point estimates. It is important to report the uncertainty of the estimates as well. Moreover, in many cases one would like to make tests between different models or data sets with respect to differences in marginal WTP. However, the preference parameters used to compute marginal WTP result from maximum likelihood estimation, and standard errors are associated with each point estimate. The problem is that one wants to find the standard deviation of an expression that is a nonlinear function of a number of parameters. There are several methods for doings this (Kling 1991; Cooper 1994). One common approach is the Delta method which involves a first-order Taylor series expansion of the WTP expression. For example, if $\mathrm{MWTP} = -\frac{\beta_i}{\lambda}$, then the variance is approximately

$$\operatorname{var}(\mathrm{MWTP}) = \frac{1}{\lambda^2}\operatorname{var}(\beta_i) + \frac{\beta_i^2}{\lambda^4}\operatorname{var}(\lambda) - 2\frac{\beta_i}{\lambda^3}\operatorname{cov}(\lambda, \beta_i). \tag{5.21}$$

Most econometric programs include routines for estimating the variance of nonlinear functions of estimated parameters using the Delta method.

Another method is the so-called Krinsky-Robb method (Krinsky and Robb 1986). This method is based on a number of random draws from the asymptotic normal distribution of the parameter estimates, and the welfare measure is then calculated for each of these draws. The standard deviation or the confidence interval is then constructed based on these draws.

A third method for computing the variance of marginal WTP is "bootstrapping," where a number of new data sets are generated by resampling, with replacement, from the estimated residuals. For each of these new data sets the model is re-estimated and welfare measures are calculated. The confidence intervals or standard deviation is then constructed based on this set of welfare measures. The Krinsky-Robb method is less computationally burdensome than bootstrapping, but its success critically depends on how closely the distribution of errors and asymptotically normal distribution coincide. Kling (1991), Cooper (1994), and Chen and Cosslett (1998) find that bootstrapping and the Krinsky-Robb method give quite similar standard deviations.

By estimating marginal WTP and the variance with any of the approaches above, one can make direct tests of difference between, say, two different groups of respondents or between different experiments. However, if one would like to pool the data from two different experiments and estimate a joint model, he or she would need to be cautious. For example, one could have conducted the same type of CE in two different countries. Now he or she wants to test whether the preferences in the two countries are the same. There are, however, two factors that could differ between the two countries: the utility parameters and the scale parameter. Remember that one cannot simply compare the estimated coefficients from two sets of data because the coefficient estimates are confounded with the scale parameter. Nevertheless, it is actually possible to construct tests of both these differences using an approach proposed by Swait and Louviere (1993). They showed that it is possible to estimate the ratio of scale parameters for two different data sets. This procedure can then be used to compare different models or to pool data from different sources (Adamowicz et al. 1994; Ben-Akiva and Morikawa 1990).

5.5.3 Willingness to Pay: Alternative Specific Experiments

When evaluating WTP in an alternative specific experiment, it is important to understand that the researcher is not certain which alternative an individual would chose. Consequently, WTP in alternative specific experiments is based on the probability of choosing the various labeled alternatives, and this is accomplished using the so-called log-sum formula (Hanemann 1999; Morey 1999), which is an expected utility version of the welfare measures used in contingent valuation. Again assume a simple linear utility function for Alternative i and respondent k:

$$v_{ik} = \beta Z_i + \lambda(y_k - p_i) + \varepsilon_{ik}. \tag{5.22}$$

Suppose a researcher wishes to estimate the respondent's WTP for a change in the attribute vector from initial conditions (Z_0) to altered conditions (Z_1). The compensating variation (CV) associated with this change is obtained by solving the equality

$$V_k\left(Z^0, p^0, y_k\right) = V_k\left(Z^1, p^1, y_k - \mathrm{CV}\right), \tag{5.23}$$

where V is the unconditional utility function. Recall that in the RUM model, respondents are assumed to choose the alternative that maximizes their utility: $\max(V_k) = \max(v_{ik} + \varepsilon_{ik}) \forall i$. This expression can be written in alternative form as

$$V_k[Z, p, y_k] = \lambda y_k + \max[\beta Z_1 - \lambda p_1 + \varepsilon_1, \ldots, \beta Z_N - \lambda p_N + \varepsilon_N], \tag{5.24}$$

where N is the number of alternatives. Inserting this expression into Eq. (5.23) for the compensating variation, we obtain the following equality:

$$\begin{aligned}&\lambda y_k + \max\left[\beta Z_1^0 - \lambda p_1^0 + \varepsilon_1^0, \ldots, \beta Z_N^0 - \lambda p_N^0 + \varepsilon_N^0\right] \\ &\quad = \lambda(y_k - \mathrm{CV}) + \max\left[\beta Z_1^1 - \lambda p_1^1 + \varepsilon_1^1, \ldots, \beta Z_N^1 - \lambda p_N^1 + \varepsilon_N^1\right].\end{aligned} \tag{5.25}$$

Now, solve Eq. (5.25) for the compensating variation:

$$\begin{aligned}\mathrm{CV} = \frac{1}{\lambda}\big\{&\max\left[\beta Z_1^0 - \lambda p_1^0 + \varepsilon_1^0, \ldots, \beta Z_N^0 - \lambda p_N^0 + \varepsilon_N^0\right] \\ &-\max\left[\beta Z_1^1 - \lambda p_1^1 + \varepsilon_1^1, \ldots, \beta Z_N^1 - \lambda p_N^1 + \varepsilon_N^1\right]\big\}.\end{aligned} \tag{5.26}$$

The final step requires expressions for the expected value of the maximum indirect utility of the alternatives. In order to do this, an assumption about the error terms in Eq. (5.22) is needed. It turns out that if the errors have an extreme value Type 1 distribution (as generally assumed for RUM models), then the expected value of the maximum utility is the so-called log-sum (or inclusive) value. For example, the log-sum value for Alternative i under initial attribute conditions can be written as

$$\max\left[\beta Z_1^0 - \lambda p_1^0 + \varepsilon_1^0, \ldots, \beta Z_N^0 - \lambda p_N^0 + \varepsilon_N^0\right] = \ln \sum_{i=1}^{N} \mathrm{e}^{v_i^0}. \tag{5.27}$$

This result leads to a very convenient expression for computing the compensating variation:

$$\mathrm{CV} = \frac{1}{\lambda}\left\{\ln \sum_{i=1}^{N} \mathrm{e}^{v_i^0} - \ln \sum_{i=1}^{N} \mathrm{e}^{v_i^1}\right\}, \tag{5.28}$$

which is simply the difference between the expected values of maximum utility for the initial and altered attribute levels for Alternative i, divided by the marginal utility of money.

5.6 Relaxing the Assumptions of the Conditional Logit Model

Up to this point, two assumptions have been made to simplify the econometric analysis of the conditional logit model. First, it was assumed that everyone in the population has the same preference structure. This assumption restricts the β's to be the same for all members of the population. Second, it was assumed that the ratio of choice probabilities between any two alternatives was unaffected by other alternatives in the choice set. This property (independence of irrelevant alternatives) results in limited substitution possibilities.

This section looks at a few models that relax these assumptions. In particular, it will focus on models that relax the assumption of identical preference parameters for all respondents, and it will look at three modifications: (1) including interaction effects, (2) estimating a latent class/finite mixture model, and (3) using a random parameter/mixed logit approach. Regarding the independence of irrelevant alternatives property, the main approach to address this issue has been the nested logit model (Ben-Akiva and Lerman 1985; Louviere et al. 2000).

5.6.1 Interaction Effects

Individual- (respondent-) specific variables (age, wealth, etc.) cannot be examined directly in a conditional logit model because these variables do not vary across alternatives. Thus, individual-specific variables drop out of the utility difference. However, individual-specific variables can interact with alternative specific attributes to provide some identification of attribute parameter differences in response to changes in individual characteristics. For example, interacting age with the price attribute would generate information on the marginal utility of money (price) as a function of age. This is a simple approach that provides insight into heterogeneity of consumers, but it assumes we already know the elements that lead to heterogeneity (those items included as interaction effects) and results in many parameters and potential collinearity problems.

5.6.2 Latent Class/Finite Mixture Model

A more advanced approach is to use a latent class/finite mixture model in which it is assumed that respondents belong to different preference classes that are defined by a small number of segments. Suppose S segments exist in the population, each with different preference structures and that individual k belongs to segment s (s = 1, …, S). The conditional indirect utility function can now be expressed as $V_{ik|s} = v_{ik|s} + \varepsilon_{ik|s}$. For simplicity, one can write the deterministic part of utility as $v_{ik} = \beta Z_i$, where again Z_i is a vector of attributes that now includes the monetary attribute. The preference parameters (β) vary by segment, so that one can write the indirect utility function as $V_{ik|s} = \beta_s Z_i + \varepsilon_{ik|s}$. The probability of choosing Alternative i depends on the segment one belongs to and can be expressed as

$$P_{ik|s} = \frac{\exp(\beta_s Z_i)}{\sum_{j=1}^{N} \exp(\beta_s Z_k)}, \tag{5.29}$$

where the β's are segment-specific utility parameters (and scale is fixed at 1).

Now let there be a process describing the probability of being included in a particular segment as a function of demographic (and other) information. Following Boxall and Adamowicz (2002), Swait (1994), and Gupta and Chintagunta (1994), that process can be specified as a separate logit model to identify segment membership as

$$P_{ks} = \frac{\exp(\delta_s X_i)}{\sum_{j=1}^{N} \exp(\beta_s Z_k)}, \tag{5.30}$$

where X is a set of individual characteristics and delta is a vector of parameters.

Let P_{iks} be the joint probability that individual k belongs to segment s and chooses Alternative i. This is also the product of the probabilities defined in Eqs. (5.29) and (5.30): $P_{iks} = P_{ik}|s \times P_{ks}$. The probability that individual k chooses i becomes the key component in the finite mixture or latent class approach:

$$P_{ik} = \sum_{s=1}^{S} P_{ik|s} P_{ks} = \sum_{s=1}^{S} \frac{\exp(\beta_s Z_i)}{\sum_{j=1}^{N} \exp(\beta_s Z_k)} \frac{\exp(\delta_s X_i)}{\sum_{j=1}^{N} \exp(\beta_s Z_k)}. \tag{5.31}$$

The joint distribution of choice probability and segment membership probability is specified and estimated in this model. Note that this approach provides information on factors that affect or result in preference differences. That is, the parameters in the segment membership function indicate how the probability of being in a specific segment is affected by age, wealth, or other elements included in the segment membership function. Further details on this approach to heterogeneity can be found in Swait (1994), Boxall and Adamowicz (2002), or Shonkwiler and Shaw (1997).

Note that the ratio of probabilities of selecting any two alternatives would contain arguments that include the systematic utilities of other alternatives in the choice set. This is the result of the probabilistic nature of membership in the elements of S. The implication of this result is that independence of irrelevant alternatives need not be assumed (Shonkwiler and Shaw 1997).

One issue with latent class models is the choice of number of classes, S. The determination of the number of classes is not part of the maximization problem, and it is not possible to use conventional specification tests such as a likelihood ratio tests. Some sort of information criteria are sometimes used (Scarpa and Thiene 2005), as well as stability of the parameters in the segments as tools to assess the best number of classes to represent the data.

5.6.3 Random Parameter/Mixed Logit Model

Another advanced approach to identifying preference heterogeneity is based on the assumption that parameters are randomly distributed in the population. Then, the heterogeneity in the sample can be captured by estimating the mean and variance of the random parameter distributions. This approach is referred to as random parameter logit or mixed logit modeling (Train 1998). In order to illustrate the random parameter logit model one can write the utility function of Alternative i for individual k as

$$v_{ik} = \beta Z_i + \varepsilon_{ik} = \bar{\beta} Z_i + \tilde{\beta}_k Z_i + \varepsilon_{ik}, \tag{5.32}$$

where, again, Z_i is a vector of attributes, including the monetary attribute. With this specification, the parameters are not fixed coefficients, but rather they are random. Each individual's coefficient vector, β, is the sum of the population mean, $\bar{\beta}$, and an individual deviation, $\tilde{\beta}_k$. The stochastic part of utility, $\tilde{\beta}_k Z_i + \varepsilon_{ik}$, is correlated among alternatives, which means that the model does not exhibit the independence of . It is assumed that the error terms are independently and identically distributed Type I extreme value.

Assume that the coefficients β vary in the population with a density distribution $f(\beta|\theta)$, where θ is a vector of the underlying parameters of the taste distribution. The probability of choosing Alternative i depends on the preferences (coefficients). The conditional probability of choosing Alternative i is

$$P_{ik|\beta} = \frac{\exp(\beta Z_i)}{\sum_{j=1}^{N} \exp(\beta Z_k)}. \tag{5.33}$$

Following Train (1998), the unconditional probability of choosing Alternative i for individual k can then be expressed as the integral of the conditional probability in (5.33) over all values of β:

$$P_{ik|\theta} = \int P_{ik|\beta} f(\beta|\theta) \mathrm{d}\beta = \int \frac{\exp(\beta Z_i)}{\sum_{j=1}^{N} \exp(\beta Z_k)} f(\beta|\theta) \mathrm{d}\beta. \tag{5.34}$$

In general, the integrals in Eq. (5.34) cannot be evaluated analytically, so one has to rely on simulation methods (Train 2003).

It is important to point out the similarities between the latent class model and the random parameter logit model. The probability expression (Eqs. 5.31 and 5.34) are both essentially weighted conditional logit models. Equation (5.31) reflects a finite weighting or mixture, whereas Eq. (5.34) is a continuous mixture.

The random parameter logit model requires an assumption to be made regarding the distribution of the coefficients. Note that it is not necessary for all parameters to follow the same distribution, and not all parameters need to be randomly distributed. The choice of distribution is not a straightforward task. In principle, any distribution could be used, but in previous applications the most common ones have been the normal and the log-normal distribution. Other distributions that have been applied are the uniform, triangular, and Raleigh distributions.

There are several aspects that one could consider when determining the distribution of the random parameters. First, one might want to impose certain restrictions. The most natural one might be that all respondents should have the same sign for the coefficients. Of the previously discussed distributions, only the log-normal distribution has this property. For example, if one assumes that the cost coefficient is log-normally distributed, it ensures that all individuals have a nonpositive price coefficient. In this case, the log-normal coefficients have the following form:

$$\beta_k = \pm\exp(b_k + \eta_k), \tag{5.35}$$

where the sign of coefficient β_k is determined by the researcher according to expectations, b_k is constant and the same for all individuals, and η_k is normally distributed across individuals with mean and variance equal to 0 and σ_k^2, respectively. This causes the coefficient to have the following properties: (1) median $= \exp(b_k)$; (2) mean $= \exp\left(b_k + \sigma_k^2/2\right)$; and (3) standard dev $= \exp\left(b_k + \sigma_k^2/2\right)\left(\exp\left(\sigma_k^2\right) - 1\right)^{0.5}$. While the log-normal distribution seems like a reasonable assumption, there may be some practical problems in its use. First, experience has shown that this distribution often causes difficulty with convergence in model estimation, likely because of the restriction it places that all respondents have the same sign on the associated coefficient. Another problem with the log-normal distribution is that the estimated welfare measures could be extremely high because values of the cost attribute close to zero are possible.

5.7 Application to Swedish Wetlands

To illustrate how the econometric models that have been described can be interpreted and inform policy decisions, an empirical example based on data collected in a mail survey regarding Swedish citizens' valuation of wetland attributes is presented. For purposes of this chapter, the following example is modified from data descriptions and analyses presented elsewhere (Carlsson et al. 2003). This example is used because it illustrates various econometric specifications in a concise way. However, it is acknowledged that the attribute specifications do not reflect current best practices regarding utility endpoints, and if the experiment had been conducted today, the specifications would have been improved (especially regarding the design of the biodiversity attribute). Given this caveat, the example is first used to illustrate the basic multinomial logit model, and then it is extended to the more advanced econometric models.

In Sweden and elsewhere, there is an increasing interest in the restoration and construction of wetlands. The purpose of the choice experiment was to identify public preferences for the characteristics of a wetland area located in southern Sweden. The attributes and their levels are presented in Table 5.7.

Table 5.7 Attributes and attribute levels for a wetland choice experiment

Attribute	Description	Levels
Cost	The lump-sum cost for the individual if the alternative is chosen	SEK[a] 200, 400, 700, 850
Landscape vegetation	The land cover type surrounding the wetland	Forest, meadow
Biodiversity	Alternative levels of plant, animal, and insect species	Low, medium, high
Sport fish	Improved habitat for sport fish such as bass and pike	No, yes
Fence	The wetland is enclosed with a 1-m fence in order to prevent drowning accidents	No, yes
Crayfish	Swedish crayfish are established and harvesting is allowed	No, yes
Walking trails	Walking trails are constructed with signs describing the plant and animal life	No, yes

[a]One Swedish krona (SEK) = approximately $0.15

Wetland attributes	**Alternative 1**	**Alternative 2**	**Alternative 3**
Landscape vegetation	Forest	Forest	Meadow
Biodiversity	Low	Low	High
Sport fish	No	Yes	No
Fence	No	No	Yes
Crayfish	No	Yes	No
Walking trails	No	No	Yes
Cost	SEK 0	SEK 850	SEK 400
I would choose: (please check one box)	☐	☐	☐

Fig. 5.1 Example of a choice set for a wetland choice experiment

In the survey, respondents made selections from four choices sets, and in each set they had to choose between three alternatives, one of which was a base alternative (Alternative 1, simple ponds) with no improvements and low biodiversity. An example choice set is presented in Fig. 5.1.

The results are presented in sequence, beginning with the multinomial logit model and then moving on to the extensions.

5.7.1 Results from the Multinomial Logit Model

The results from the maximum likelihood estimation of a multinomial logit model are shown in Table 5.8; note that not all of the attribute coefficients are statistically significant.[22] For the attribute "landscape vegetation," the dummy variable for meadowland is negative but statistically insignificant, which indicates that there is no significant difference in preferences between meadowland and forest vegetation. For the biodiversity attribute, the dummy variables for medium and high biodiversity are both positive and statistically significant. This indicates that respondents prefer both medium and high levels of biodiversity relative to the baseline of low biodiversity. The dummy variables for improved sport fish habitat and walking trails are both positive and statistically significant, which indicates that respondents prefer having these attributes relative to a baseline of no sport fish habitat improvement and no walking trails, respectively. The dummy variables for crayfish and fence construction are both negative and statistically significant, which indicates that respondents dislike establishment of crayfish and construction of a fence around the area. The coefficient of the total cost attribute is negative and significant, which, as expected, indicates that respondents have a positive marginal utility of money. Finally, the coefficient of the alternative specific constant is negative and insignificant. Recall that, as defined here, the alternative specific constant represents the utility of choosing the status quo alternative (Alternative 1). This result suggests the absence of status quo bias (choosing the status quo regardless of attribute levels) and that respondents made choices strictly based on the level of the attributes.

Marginal WTP values are estimated by dividing the attribute coefficient by the marginal utility of money.[23] The marginal WTP values indicate the strength of respondent preferences for the attributes expressed in Swedish kronor. For example, on average, respondents are willing to pay almost 680 kronor (roughly $102) in a lump sum to obtain a high biodiversity wetland compared with a low-biodiversity wetland. The marginal WTP to obtain a medium biodiversity wetland is 512 kronor. A simple *t*-test (using standard errors calculated using the Delta method) reveals that there is no statistically significant difference between respondents willingness to pay for high relative to medium biodiversity. Marginal WTP for improved sport fish habitat is 351 kronor relative to the base level of no action. The marginal WTP for the construction of a fence is −169 kronor. This indicates the marginal WTP is 169 kronor less than the marginal WTP for the base level, which was no fenced area.

[22]In all of the tables, *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, and * denotes significance at the 0.10 level. Also, standard errors (s.e.) of the coefficients are shown in parentheses below the coefficients.

[23]The coefficients shown in Model 1 (Table 5.8) have been rounded to three decimal places. However, the marginal WTP values shown in Table 5.8 were computed before rounding. Computation of marginal WTP values based on the coefficients shown in Model 1 will therefore result in slightly different values than reported in Table 5.8.

Table 5.8 Multinomial logit parameter estimates for the wetland choice experiment

	Model 1 coefficient (s.e.)	**Model 1 marginal WTP (s.e.)**
ASC (Alternative specific constant for status quo alternative)	−0.119 (0.125)	–
Landscape vegetation is meadowland	-0.052 (0.061)	-45 (53)
High biodiversity	0.784*** (0.086)	679*** (106)
Medium biodiversity	0.591*** (0.085)	512*** (94)
Sport fish, improved habitat	0.405*** (0.062)	351*** (65)
Fence constructed	−0.194*** (0.062)	−169*** (58)
Crayfish established	−0.130** (0.061)	−113** (55)
Walking trails constructed	0.753*** (0.063)	653*** (92)
Cost	−0.001*** (0.0001)	–

5.7.2 *Results from Model Extensions*

This section examines the results from models that relax the standard multinomial logit assumption, starting with a model using interaction terms (Table 5.9). This approach will help develop intuition regarding how key variables in the latent class and random parameter logit models might be identified. The results of two models are reported—one with interaction between one socio-economic characteristic—male (a dummy variable equal to one if the respondent is a male)—and the alternative specific constant (Model 2), and another model that interacts this characteristic with the alternative specific constants and all of the other attributes except cost (Model 3).

This section will not comment on all of the results in detail because the interpretation of the coefficient estimates and WTP values have been discussed already. The one aspect that is important to point out is the result in Model 2, where the alternative specific constant is interacted with the male dummy variable. In Model 1, without interaction terms, the alternative specific constant is negative but statistically insignificant. In Model 2, with an interaction term, there are significant differences between males and females. Females have a negative alternative specific constants (−0.275), which is statistically significant. The negative sign indicates that choosing the status quo decreases indirect utility for this group of respondents (choosing alternatives to the status quo increases indirect utility). The interaction

Table 5.9 Multinomial logit parameter estimates for the wetland example with interaction terms

	Model 2 coefficient (s.e.)	**Model 3 coefficient (s.e.)**
ASC (for status quo alternative)	−0.275** (0.138)	−0.001 (0.161)
ASC × male	0.317*** (0.114)	0.238 (0.224)
Landscape vegetation is meadowland	−0.051 (0.061)	0.219*** (0.085)
Landscape vegetation is meadowland × male	–	−0.580*** (0.125)
High biodiversity	0.784*** (0.086)	1.001*** (0.122)
High biodiversity × male	–	−0.418** (0.174)
Medium biodiversity	0.591*** (0.085)	0.737*** (0.118)
Medium biodiversity × male	–	−0.271 (0.172)
Sport fish, improved habitat	0.406*** (0.062)	0.340*** (0.084)
Sport fish, improved habitat × male	–	0.150 (0.125)
Fence constructed	−0.195*** (0.062)	−0.095 (0.084)
Fence constructed × male	–	0.234* (0.125)
Crayfish established	−0.131** (0.061)	−0.125 (0.084)
Crayfish established × male	–	−0.003 (0.125)
Walking trails constructed	0.753*** (0.063)	0.777*** (0.084)
Walking trails constructed × male	–	−0.005 (0.128)
Cost	−0.001*** (0.0001)	0.001*** (0.0001)

term between the alternative specific constant and the male dummy variable is positive. The alternative specific constant for males is the sum of the two terms (i.e., −0.275 + 0.317 = 0.042), which is positive.

When estimating marginal WTP values for the attributes in Model 3, note that it is possible to estimate multiple values for each attribute. First, one can estimate one value for women, which will be the ratio of the attribute coefficient to the marginal utility of money. One can estimate a second value for men, which will be the ratio of the sum of the attribute coefficient plus the interaction term to the marginal utility of money. And third, one can estimate the sample mean WTP values as well. This is calculated as a weighted sum of female and male WTP values, where the weights are the percentage of females and males in the sample (Table 5.10).

Again, this section will not discuss all the results in detail. The first observation that can be made is that the average WTP is almost the same in the two models when both males and females are included, as would be expected. However, the model with interaction terms (Model 3) reveals that, in some instances, there are large differences between men and women. For example, females have a positive WTP for meadowland, while males have a negative WTP. Furthermore, females have a higher WTP for high and medium levels of biodiversity (compared with low levels) relative to males.

The results obtained using a random parameter logit model and a latent class model are shown in Table 5.11. In the random parameter logit model, all the nonmonetary attributes were specified as being normally distributed. In the latent class model, two classes were specified and the gender of the respondent was used to explain class membership.

For the random parameter logit model, there are two coefficients estimated for each of the random parameters, where the first is an estimate of the mean preference and the second is an estimate of the standard deviation of preferences across the sample. Three of the standard deviation parameters are statistically significant, which indicates that the model is capturing unobserved heterogeneity. Furthermore, the standard deviation estimates are large relative to the mean values and indicate that respondents had reverse preferences (opposite signs) for some attributes. For example, the coefficient for the fence attribute is −0.494, and the standard deviation is 1.605. This indicates that, on average, respondents disliked fencing in the

Table 5.10 Marginal WTP values (and standard errors) comparing the basic multinomial logit model results and a multinomial logit model with interaction terms

	Model 1	**Model 3**		
	Sample	**Female**	**Male**	**Sample**
Landscape vegetation is meadowland	−45 (53)	183** (72)	−301*** (83)	−47 (52)
High biodiversity	679*** (106)	836*** (134)	487*** (117)	671*** (102)
Medium biodiversity	512*** (94)	616*** (120)	389*** (114)	507*** (92)
Sport fish, improved habitat	351*** (65)	284*** (76)	410*** (88)	344*** (63)
Fence constructed	−169*** (58)	−79 (71)	−275*** (85)	−172*** (56)
Crayfish established	−113** (55)	−105 (72)	−102 (78)	−103* (54)
Walking trails constructed	653*** (92)	649*** (102)	645*** (105)	647*** (89)

Table 5.11 Parameter estimates of the random parameter and latent class logit models for the wetland choice experiment example

	Random parameter logit model		Latent class logit model	
	Coefficient (s.e.)	Coefficient (s.e.)	Class 1 (s.e.)	Class 2 (s.e.)
ASC (for status quo alternative)	0.048 (0.211)	–	−0.328 (0.324)	0.039 (0.381)
Landscape vegetation is meadowland	−0.220 (0.158)	1.489** (0.695)	−1.105*** (0.413)	0.782*** (0.288)
High biodiversity	1.402*** (0.309)	1.019 (0.944)	0.313 (0.276)	1.145*** (0.286)
Medium biodiversity	1.031*** (0.242)	0.859 (1.025)	0.136 (0.265)	1.127*** (0.269)
Sport fish, improved habitat	0.788*** (0.202)	0.689 (1.008)	0.913*** (0.221)	0.131 (0.188)
Fence constructed	−0.494*** (0.186)	1.605** (0.663)	−0.586*** (0.216)	0.021 (0.166)
Crayfish established	−0.477*** (0.218)	1.982*** (0.682)	0.297 (0.196)	−0.095 (0.159)
Walking trails constructed	1.472*** (0.362)	0.383 (0.662)	1.087*** (0.262)	0.839*** (0.177)
Cost	−0.002*** (0.0006)		−0.002*** (0.0004)	−0.001*** (0.0004)
Latent class probability model				
Constant			−0.595 (0.527)	–
Male			1.359*** (0.290)	–
Average class probabilities			0.510	0.490

wetland, but that a fraction of respondents had a positive preference for constructing a fence.

The latent class model was estimated with two classes (male and female), and one can see some distinct differences between the two classes. For example, in the first class, there are no statistically significant preferences for improvement in biodiversity, while in the second class, biodiversity is the most important attribute. In the class inclusion equation, the gender of the respondent is statistically significant, which means that it is more likely that a man belongs to Class 1 than to Class 2. Not surprisingly, these results are consistent with results reported in Model 3 (Table 5.9).

Next the marginal WTP values for the random parameter logit and latent class logit models are estimated (Table 5.12).

For the random parameter logit model, the point estimates of marginal WTP are found to be mostly similar to the values obtained from the standard logit model

Table 5.12 Marginal WTP values from random parameter logit and latent class logit models

		Latent class logit	
	Random parameter logit (s.e.)	**Class 1 (s.e.)**	**Class 2 (s.e.)**
Landscape vegetation is meadowland	−99 (65)	−724** (293)	565** (230)
High biodiversity	634*** (101)	205 (200)	1045*** (310)
Medium biodiversity	466*** (92)	89 (177)	813*** (268)
Sport fish, improved habitat	356*** (68)	598*** (194)	95 (139)
Fence constructed	−223*** (75)	−383** (165)	15 (120)
Crayfish established	−215*** (79)	−194 (144)	−69 (118)
Walking trails constructed	665*** (91)	712*** (238)	606*** (193)

(Model 1). However, it should be noted that the estimated unobserved heterogeneity has not been taken into account (i.e., there is an estimated standard deviation of WTP values as well). For the latent class logit model, the marginal WTP values are estimated for the two classes, and as can be seen, there are considerable differences between the two classes. Again, in the second class there is a considerable WTP for biodiversity, while in the first class priority is put on fishing and walking trails.

Note two important considerations. First, for the latent class logit model, mean WTP for the sample can be estimated by taking the weighted average of the two classes, where the weights are the average class probabilities in Table 5.11. By doing that, the estimated WTP values will be similar to the WTP values of the conditional logit model and the random parameter logit model. Thus, neither the random parameter logit nor latent class logit model results in very different overall WTP values, but they do provide much richer information about the distribution of preferences over the sample.

Second, the results of the latent class logit model are consistent with what was found in the conditional logit model where the attributes were interacted with the male dummy variable (Model 3). In that model, for example, it was found that men care less about biodiversity and more about fish, while women care more about biodiversity.

5.8 Conclusions

Choice experiments have emerged as the preferred method (relative to rankings and ratings) for conducting stated-preference studies when a good or service is best characterized by a suite of attributes. This result is primarily explained by the fact that CEs mimic actual market behavior in a policy context where various dimensions of the policy attributes are under consideration. The use of CEs provides an opportunity to evaluate the welfare effects of multiple policy options and to calibrate ex ante value estimates to ex post policies. The realistic nature of well-designed CEs is complemented by the power of random utility maximization models to describe decision-making involving trade-offs among attributes and by recent advances that integrate RUM models with experimental design. This evolution of thinking and practice provides powerful tools for collecting and analyzing data that can assist the estimation of environmental (and other) values in a benefit-cost context and can assist benefit transfers when funds for original studies are scarce.

Despite the mathematical and statistical nature of many of the topics presented in this chapter, the authors want to highlight the importance of the preliminary steps in designing a CE, which rely more on developing communication skills among economists, other scientists, policymakers, and members of the public. A CE should only be used for policy analysis when it is clearly the best method available for data collection and analysis. The ability to identify which attributes of a policy issue people really care about and how, where, and when environmental changes may impact use and passive-use values are topics that require careful deliberation and meaningful conversations in focus group settings. Also, as emphasized in the description of experimental design procedures, conducting meaningful survey pretests is essential for constructing an efficient experimental design. Without due care and attention to these steps, CE applications will not provide the desired information. However, by following the guidance provided in this chapter, researchers will be able to design and analyze a CE that provides credible value estimates that can provide meaningful support for decision-making.

Appendix 1: Choice Experiments and Behavior

Many choice experiments that appear in the literature examine passive-use values or "total economic values" in that they ask respondents to choose between options that may affect outcomes associated with passive-use values (e.g., endangered species) or use values (recreation enjoyment, etc.). These are often framed as referenda or social choices. The case study examined in this chapter is an example of this type of choice experiment. However, choice experiments can also be based on behavioral choices alone.

In other literature, such as transportation and marketing, choice experiments are typically used to assess how behavior such as transport mode choice or choice of a product will vary with different attributes. Indeed, the earliest applications of choice experiments in economics included cases of recreation site choice or property choices. There are a variety of reasons to use choice experiments in the analysis of behavior, even if revealed preference data on such choices are available. Choice experiments can present attributes that are outside the range of the existing set of attributes (e.g., higher levels of fishing catch rates or higher levels of congestion on hiking trails) and reflect new attributes (e.g., environmental labels on consumer products), and experimental design can help in identifying parameters on attributes that are typically correlated in the real world (e.g., water quality and fish catch). Choice experiment data can also be combined with revealed preference data to help calibrate stated-preference responses or to compare stated and revealed preference information. Efforts in “data fusion” or the combination of stated and revealed preference information have included analyses of recreation, property choices, as well as the integration of perceived and objective measures of attributes. A key aspect of the use of choice experiments in behavioral contexts is that the collection of stated choice data support a model based on economic theory such as the travel cost model of recreation choice behavior (Bockstael and McConnell 2007) or the random utility approach to property choice and valuation (Phaneuf et al. 2013). To help illustrate these approaches we provide two examples of choice experiments that involve linkages to behavioral choices—recreation site choice and property choices.

In the recreation case, the respondent is asked to choose between two moose hunting sites in which the attributes include distance (the travel cost) and other characteristics of the hunting site (Fig. 5.2). The attributes are described in a way that hunters typically view them, and in a way that they can be translated to “objective” measures of wildlife populations and site descriptions. This case study, drawn from Adamowicz et al. (1997), included information on actual site choices and elicited information from hunters on their perceptions as well as actual measures of attributes. This data set has been used in other applications including the assessment of unobserved heterogeneity in stated and revealed preference data (von Haefen and Phaneuf 2008). Note that in this case, the hunting sites are described as “generic” sites (Site A and Site B). In some contexts these can be described as actual sites with “labels” such as the name of the area or the administrative label

Assuming that the following areas were the ONLY areas available, which one would you choose on your next hunting trip, if either?

Features of hunting area	**Site A**	**Site B**	
Distance from home to hunting area	50 kilometers	150 kilometers	
Quality of road from home to hunting area	Mostly gravel or dirt, some paved	Mostly paved, some gravel or dirt	
Access within hunting area	No trails, cutlines, or seismic lines	Newer trails, cutlines or seismic liens passable with a four wheel drive vehicle	Neither Site A nor Site B
Encounters with other hunters	No hunters, other than those in my hunting party, are encountered	Other hunters, on all terrain vehicles, are encountered	I will NOT go moose hunting
Forestry activity	Some evidence of recent logging found in the area	No evidence of logging	
Moose population	Evidence of less than 1 moose per day	Evidence of 3-4 moose per day	
Check ONE and only one box	☐	☐	☐

Fig. 5.2 Example of a recreational hunting site choice

(see Boxall and Adamowicz, 2002, for an application to canoeing sites). Similar types of choice experiments have been used to elicit tradeoffs in the context of subsistence resource use (trading off distance with traditional use hunting or caloric expenditures for fuelwood collection (see Adamowicz et al., 2004).

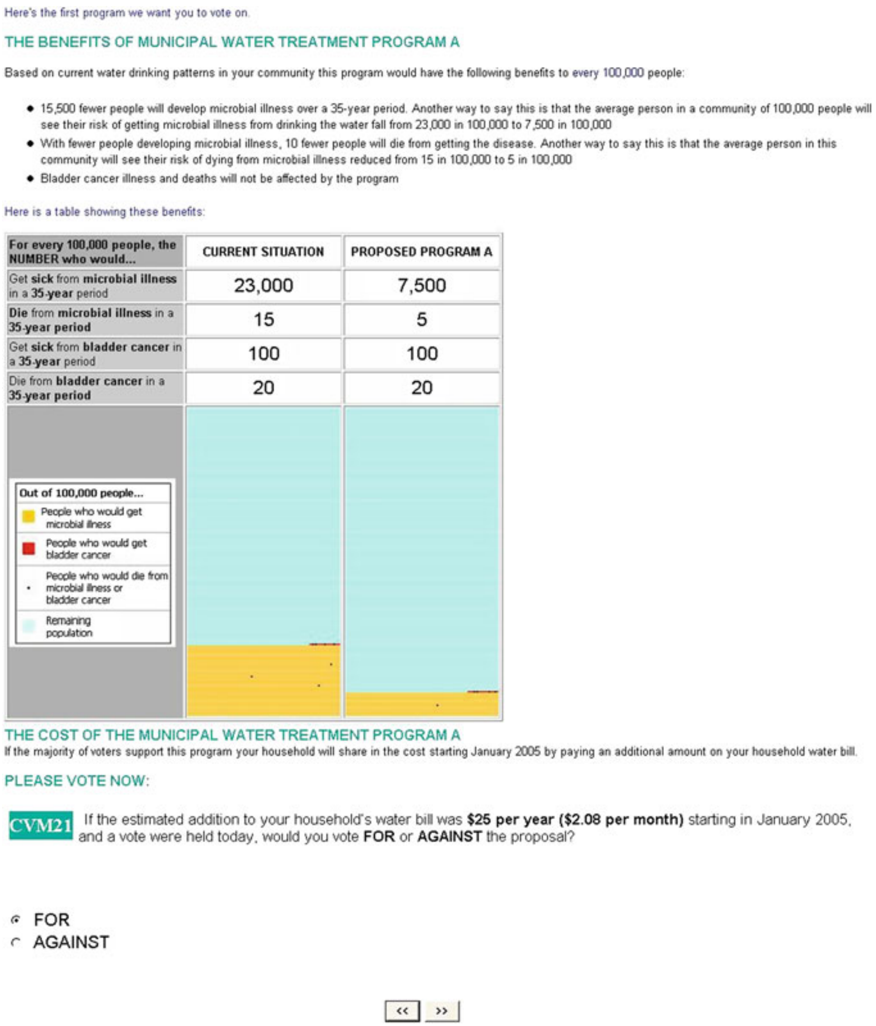
The second example (Fig. 5.3) is similar to the property choice cases used in Phaneuf et al. (2013) and is based on Kim (2014). In this case, information about a respondent's current house is elicited. This information is used as the "base" case for the choice experiment, and attributes are presented that describe changes to the house (larger area, different water quality in the adjacent lake, etc.). This choice experiment uses an experimental design referred to as a "pivot design" in that it pivots the choices around the currently held option or behavior (Hess and Rose 2009).

Appendix 2: Choice Experiments and the Value of Health Risk Reduction

A nonmarket value that is very important in policy analysis is the value of mortality risk reduction, often referred to as the value of statistical life (see Cameron, 2010, for a review of the issues and a thorough critique of the term "value of statistical life"). The value of mortality risk reductions often comprises over 80% of the monetary value of air pollution reduction policies such as the assessment of the U.S. Clean Air Act Amendments. Mortality risk values have typically been measured using hedonic wage models (see Chap. 7) wherein the impact of changing job risk characteristics are reflected in higher wages, all else held constant. Over the past few decades however stated-preference methods have been increasingly used to elicit the value of risk reductions. In a typical setting a respondent is informed about baseline risk levels and then presented with a treatment that offers a reduction in health risks, but at a cost. The tradeoff between cost and risk change provides a measure of the monetary value of risk reduction.

Characteristics	**Your current house**	**House A**	**House B**
House size		30% smaller	30% larger
House age		Newer	20 years older
Distance from the lake		More than 500 meters farther	250 meters farther
Water quality		10% worse	20% better
House price		30% more	30% less
Which house would you choose?	○	○	○

Fig. 5.3 Example of a property choice



Here's the first program we want you to vote on.

THE BENEFITS OF MUNICIPAL WATER TREATMENT PROGRAM A

Based on current water drinking patterns in your community this program would have the following benefits to every 100,000 people:

- 15,500 fewer people will develop microbial illness over a 35-year period. Another way to say this is that the average person in a community of 100,000 people will see their risk of getting microbial illness from drinking the water fall from 23,000 in 100,000 to 7,500 in 100,000
- With fewer people developing microbial illness, 10 fewer people will die from getting the disease. Another way to say this is that the average person in this community will see their risk of dying from microbial illness reduced from 15 in 100,000 to 5 in 100,000
- Bladder cancer illness and deaths will not be affected by the program

Here is a table showing these benefits:

For every 100,000 people, the NUMBER who would...	CURRENT SITUATION	PROPOSED PROGRAM A
Get **sick** from **microbial illness** in a **35-year** period	23,000	7,500
Die from **microbial illness** in a **35-year period**	15	5
Get **sick** from **bladder cancer** in a **35-year** period	100	100
Die from **bladder cancer** in a **35-year period**	20	20

THE COST OF THE MUNICIPAL WATER TREATMENT PROGRAM A

If the majority of voters support this program your household will share in the cost starting January 2005 by paying an additional amount on your household water bill.

PLEASE VOTE NOW:

CVM21 If the estimated addition to your household's water bill was **$25 per year ($2.08 per month)** starting in January 2005, and a vote were held today, would you vote **FOR** or **AGAINST** the proposal?

- FOR
- AGAINST

<< >>

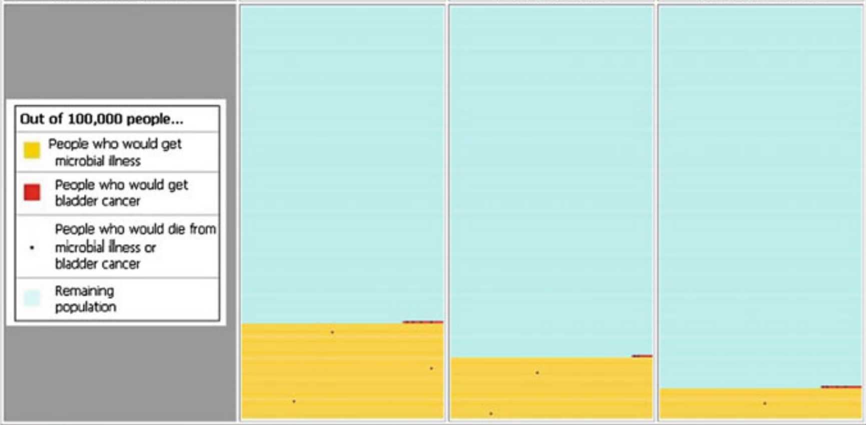
Fig. 5.4 Example of a two-alternative choice task eliciting values of health risk reductions

While contingent valuation has typically been used to measure mortality risks (e.g., Krupnick et al. 2002), choice experiments are increasingly being used to separate mortality from morbidity risks (Adamowicz et al. 2011) or to include risk context elements within the valuation tasks, such as latency (a delay in the timing of the benefits from the risk reduction), type of risk, or other elements (Alberini and Ščasný 2011).

The value of mortality risk reduction is expressed as the willingness to pay for a small reduction in the probability of mortality. For example, if individuals are willing to pay $10,000 for a 1% reduction in their risk of dying in a year, this would translate into a $1,000,000 value of statistical life (100 people valuing a 1% risk reduction would equal one "statistical life" and 100 times $10,000 is $1,000,000). A choice

This is the first scenario we want you to vote on.

For every 100,000 people, the NUMBER who would...	CURRENT SITUATION	PROPOSED PROGRAM A	PROPOSED PROGRAM B
Get **sick** from **microbial illness** in a **35-year** period	23,000	15,000	7,500
Die from **microbial illness** in a **35-year** period	15	10	5
Get **sick** from **bladder cancer** in a **35-year** period	100	50	100
Die from **bladder cancer** in a **35-year** period	20	10	20
Change to your water bill starting in January, 2005	No Change	Increase $75 per year ($6.25 per month)	Increase $75 per year ($6.25 per month)



DC1 If there were a referendum, I would vote for...

CHECK ONE ONLY

- Current Situation
- Proposed Program A
- Proposed Program B

Fig. 5.5 Example of a three-alternative choice experiment eliciting values of health risk reduction

experiment, therefore, can be designed to elicit trade-offs between a current situation (with the mortality risk presented) and an "improved" situation with the risks reduced.

The figures that follow illustrate these choices based on the case of water risks in Adamowicz et al. (2011). The respondent faces a base or status quo set of risks (mortality and morbidity from cancer and microbial impacts) and trades them off against a set of new programs. A two-alternative case (Fig. 5.4) and a three-alternative case (Fig. 5.5) are presented to illustrate that in this context, the choice experiment can be presented like a contingent valuation referendum task as well as in the context of a multiple alternative choice experiment (see, however, Zhang and Adamowicz, 2011). Note also that the risks are presented numerically (number of illnesses and deaths) as well as graphically, using grids of points to represent the risks.

Risk communication is a particularly important aspect of the assessment of health risk reductions. The random utility model that arises from these choices provides the marginal utility of risk and the marginal utility of money, and thus the value of a change in risk can be derived. Adamowicz et al. (2011) examined risks in water, while other researchers have examined climate change risks (Ščasný and Alberini 2012), risks from nuclear versus fossil fuel based energy (Itaoka et al. 2006), and mortality risks from different risk contexts, including transport, respiratory illness, and cancer (Alberini and Ščasný 2011). One of the most sophisticated choice experiments examining the value of health risk reductions, from Cameron and DeShazo (2013), examined latency, timing of illness, type of illness, and other factors that affect health.

References

Adamowicz, W., Boxall, P., Haener, M., Zhang, Y., Dosman, D. & Marois, J. (2004). An assessment of the impacts of forest management on aboriginal hunters: Evidence from stated and revealed preference data. Forest Science, 50, 139-152.

Adamowicz, W., Dupont, D., Krupnik, A. & Zhang, J. (2011). Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. Journal of Environmental Economics and Management, 61, 213-226.

Adamowicz, W., Louviere, J. & Swait, J. (1998). Introduction to attribute-based stated choice methods. Final report to the Resource Valuation Branch, Damage Assessment Center, National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Edmonton, AB, Canada: Advantis.

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining stated and revealed preference methods for valuing environmental amenities. Journal of Environmental Economics and Management, 26, 271-292.

Adamowicz, W., Swait, J., Boxall, P., Louviere, J. & Williams, M. (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Journal of Environmental Economics and Management, 32, 65-84.

Alberini, A. & Ščasný, M. (2011). Context and the VSL: Evidence from a stated preference study in Italy and the Czech Republic. Environmental and Resources Economics, 49, 511-538.

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. Psychological Review, 77, 153-170.

Ben-Akiva, M. & Lerman, S. R. (1985). Discrete choice analysis: Theory and application to travel demand. Cambridge, MA: MIT Press.

Ben-Akiva, M. & Morikawa, T. (1990). Estimation of travel demand models from multiple data sources. In M. Koshi (Ed.), Transportation and traffic theory (pp. 461-476). New York: Elsevier.

Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. Transportation Research Part B: Methodological, 29, 471-483.

Bhat, C. R. (1997). Covariance heterogeneity in nested logit models: Econometric structure and application to intercity travel. Transportation Research Part B: Methodological, 31, 11-21.

Bliemer, M. C. J. & Rose, J. M. (2011). Experimental design influences on stated choice outputs: an empirical study in air travel choice. Transportation Research Part A: Policy and Practice, 45, 63-79.

Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N. & Freeman, P. (2008). Eliciting willingness to pay without bias: Evidence from a field experiment. Economic Journal, 118, 114-137.

Bockstael, N. E. & McConnell, K. E. (2007). Environmental and resource valuation with revealed preferences: A Theoretical Guide to Empirical Models. Dordrecht, The Netherlands: Springer.

Boxall, P. C. & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: A latent class approach. Environmental and Resource Economics, 23, 421-446.

Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M. & Louviere, J. (1996). A comparison of stated preference methods for environmental values. Ecological Economics, 18, 243-253.
Boyd, J. & Krupnick, A. (2009). The definition and choice of environmental commodities for nonmarket valuation. Resources for the Future Discussion Paper 09-35. Washington, DC: Resources for the Future.
Boyle, K. J., Johnson, F. R., McCollum, D. W., Desvousges, W. H., Dunford, R. W. & Hudson, S. P. (1996). Valuing public goods: Discrete versus continuous contingent-valuation responses. Land Economics, 72, 381-396.
Bunch, D. S., Louviere, J. J. & Anderson, D. (1996). A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. Working paper. University of California-Davis.
Cameron, T. A. (2010). Euthanizing the value of a statistical life. Review of Environmental Economics and Policy, 4, 161-178.
Cameron, T. A. & DeShazo, J. R. (2013). Demand for health risk reductions. Journal of Environmental Economics and Management, 65, 87-109.
Carlsson, F., Frykblom, P. & Liljenstolpe, C. (2003). Valuing wetland attributes: An application of choice experiments. Ecological Economics, 47, 95-103.
Carlsson, F. & Martinsson, P. (2003). Design techniques for stated preference methods in health economics. Health Economics, 12, 281-294.
Carson, R. T. & Groves, T. (2007). Incentive and information properties of preference questions. Environmental and Resource Economics, 37, 181-210.
Carson, R. T. & Groves, T. (2011). Incentive and information properties of preference questions: Commentary and extensions. In J. Bennett (Ed.), The International handbook of non-market environmental valuation (pp. 300-321). Northampton, MA: Edward Elgar.
Carson, R. T. & Louviere, J. J. (2011). A common nomenclature for stated preference elicitation approaches. Environmental and Resource Economics, 49, 539-559.
Cattin, P. & Wittink, D. R. (1982). Commercial use of conjoint analysis: A survey. Journal of Marketing, 46 (3), 44-53.
Chen, H. Z. & Cosslett, S. R. (1998). Environmental quality preference and benefit estimation in multinomial probit models: A simulation approach. American Journal of Agricultural Economics, 80, 512-520.
Cooper, J. C. (1994). A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous choice contingent valuation surveys. Land Economics, 70, 111-122.
Court, A. T. (1939). Hedonic price indexes with automotive examples. In The dynamics of automobile demand (pp. 99-117). New York: General Motors.
Cummings, R. & Taylor, L. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. American Economic Review, 89, 649-665.
Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., Scarpa, R. & Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management, 63, 73-91.
Dillman, D. A. (1978). Mail and telephone surveys: The total design method. New York: Wiley.
Ferrini, S. & Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. Journal of Environmental Economics and Management, 53, 342-363.
Gan, C. E. C. & Luzar, E. J. (1993). A conjoint analysis of waterfowl hunting in Louisiana. Journal of Agricultural and Applied Economics, 25, 36-45.
Green, P. E. & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. Journal of Marketing Research, 8, 355-363.
Green, P. E. & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. Journal of Consumer Research, 5, 103-123.
Green, P. E. & Wind, Y. (1975). New way to measure consumers' judgments. Harvard Business Review, 53, 107-117.
Greene, W. H. (2002). Econometric analysis (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Griliches, Z. (1971). Hedonic price indexes for automobiles: An econometric analysis of quality change. In Z. Griliches (Ed.), Price indexes and quality change: Studies in new methods of measurement (pp. 55-77). Cambridge, MA: Harvard University Press.
Gupta, S. & Chintagunta, P. K. (1994). On using demographic variables to determine segment membership in logit mixture models. Journal of Marketing Research, 31, 128-136.
Hammond, K. R. (1955). Probabilistic functioning and the clinical method. Psychological Review, 62, 255-262.
Hanemann, W. M. (1999). Welfare analysis with discrete choice models. In J. A. Herriges and C. L. Kling (Eds.), Valuing recreation and the environment: Revealed preference methods in theory and practice (pp. 33-64). Northampton MA: Edward Elgar.
Hausman, J. & McFadden, D. (1984). Specification tests for the multinomial logit model. Econometrica, 52, 1219-1240.
Hensher, D. A., Rose, J. M. & Greene, W. H. (2005). Applied choice analysis: A primer. New York: Cambridge University Press.
Hess, S. & Rose, J. M. (2009). Should reference alternatives in pivot design SC surveys be treated differently? Environmental and Resource Economics, 42, 297-317.
Holmes, T. P. & Boyle, K. J. (2005). Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Economics, 81, 114-126.
Horne, P., Boxall, P. C. & Adamowicz, W. L. (2005). Multiple-use management of forest recreation sites: A spatially explicit choice experiment. Forest Ecology and Management, 207, 189-199.
Huber, J. & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. Journal of Marketing Research, 33, 307-317.
Itaoka, K., Saito, A., Krupnick, A., Adamowicz, W. & Taniguchi, T. (2006). The effect of risk characteristics on the willingness to pay for mortality risk reductions from electric power generation. Environmental and Resource Economics, 33, 371-398.
Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B. & Bridges, J. F. P. (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. Value in Health, 16, 3-13.
Kanninen, B. J. (2002). Optimal design for multinomial choice experiments. Journal of Marketing Research, 39, 214-227.
Kessels, R., Jones, B., Goos, P. & Vandebroek, M. (2008). Recommendations on the use of Bayesian optimal design strategies for choice experiments. Quality and Reliability Engineering International, 24, 737-744.
Kim, H. N. (2014). The demonstration and capture of an ecosystem service value: Three different methodological approaches. Ph.D. thesis. Edmonton: University of Alberta.
Kling, C. L. (1991). Estimating the precision of welfare measures. Journal of Environmental Economics and Management, 21, 244-259.
Krinsky, I. & Robb, A. L. (1986). On approximating the statistical properties of elasticities. Review of Economics and Statistics, 68, 715-719.
Krupnick, A., Alberini, A., Cropper, M., Simon, N., O'Brien, B., Goeree, R. & Heintzelman, M. (2002). Age, health, and the willingness to pay for mortality risk reductions: A contingent valuation survey of Ontario residents. Journal of Risk and Uncertainty, 24, 161-186.
Lancaster, K. J. (1966). A new approach to consumer theory. Journal of Political Economy, 74, 132-157.
Lareau, T. J. & Rae, D. A. (1989). Valuing WTP for diesel odor reductions: An application of contingent ranking technique. Southern Economic Journal, 55, 728-742.
Lindhjem, H. & Navrud, S. (2011). Using Internet in stated preference surveys: A review and comparison of survey modes. International Review of Environmental and Resource Economics, 5, 309-351.
List, J. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. American Economic Review, 91, 1498-1507.
Louviere, J. J. (1988a). Analyzing decision making: Metric conjoint analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-067. Newbury Park, CA: Sage.

Louviere, J. J. (1988b). Conjoint analysis modeling of stated preferences: a review of theory, methods, recent developments and external validity. Journal of Transport Economics and Policy, 10, 93-119.

Louviere, J. J. & Hensher, D. A. (1983). Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. Journal of Consumer Research, 10, 348-361.

Louviere, J. J. & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. Journal of Marketing Research, 20, 350-367.

Louviere, J. J., Flynn, T. N. & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. Journal of Choice Modelling, 3 (3), 57-72.

Louviere, J. J., Hensher, D. A. & Swait, J. D. (2000). Stated choice methods: Analysis and applications. Cambridge, United Kingdom: Cambridge University Press.

Louviere, J. J., Islam, T., Wasi, N., Street, D. & Burgess, L. (2008). Designing discrete choice experiments: Do optimal designs come at a price? Journal of Consumer Research, 35, 360-375.

Luce, R. D. (1959). Individual choice behavior A Theoretical Analysis. New York: Wiley.

Luce, R. D. & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology, 1, 1-27.

Mackenzie, J. (1993). A comparison of contingent preference models. American Journal of Agricultural Economics, 75, 593-603.

Manski, C. (1977). The structure of random utility models. Theory and Decision, 8, 229-254.

Marschak, J. (1960). Binary choice constraints on random utility indicators. In K. Arrow (Ed.), Mathematical methods in the social sciences, 1959: Proceedings of the first Sanford Symposium (pp. 312-329). Stanford, CA: Stanford University Press.

Mazzotta, M. J. & Opaluch, J. J. (1995). Decision making when choices are complex: A test of Heiner's hypothesis. Land Economics, 71, 500-515.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), Frontiers in econometrics (pp. 105-142). New York: Academic Press.

Morey, E. R. (1999). Two RUMs uncloaked: Nested logit models of site choice and nested logit models of participation and site choice. In J. A. Herriges and C. L. Kling (Eds.), Valuing recreation and the environment: Revealed preference methods in theory and practice (pp. 65-120). Northampton, MA: Edward Elgar.

Naidoo, R. & Adamowicz, W. L. (2005). Biodiversity and nature-based tourism at forest reserves in Uganda. Environment and Development Economics, 10, 159-178.

Phaneuf, D. J., Taylor, L. O. & Braden, J. B. (2013). Combining revealed and stated preference data to estimate preferences for residential amenities: A GMM approach. Land Economics, 89, 30-52.

Qin, P., Carlsson, F. & Xu, J. (2011). Forest tenure reform in China: A choice experiment on farmers' property rights preferences. Land Economics, 87, 473-487.

Rae, D. A. (1983). The value to visitors of improving visibility at Mesa Verde and Great Smoky National Parks. In R. D. Rowe and L. G. Chestnut (Eds.), Managing air quality and scenic resources at national parks and wilderness areas. Boulder, CO: Westview Press.

Ready, R. C., Champ, P. A. & Lawton, J. L. (2010). Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. Land Economics, 86, 363-381.

Roe, B., Boyle, K. J. & Teisl, M. F. (1996). Using conjoint analysis to derive estimates of compensating variation. Journal of Environmental Economics and Management, 31, 145-159.

Rolfe, J., Bennett, J. & Louviere, J. (2000). Choice modelling and its potential application to tropical rainforest preservation. Ecological Economics, 35, 289-302.

Rose, J. M. & Bliemer, M. C. J. (2009). Constructing efficient stated choice experimental designs. Transport Reviews: A Transnational Transdisciplinary Journal, 29, 587-617.

Sándor, Z. & Wedel, M. (2001). Designing conjoint choice experiments using manager's prior beliefs. Journal of Marketing Research, 38, 430-444.

Scarpa, R. & Rose, J. M. (2008). Design efficiency for non-market valuation with choice modelling: How to measure it, what to report and why. Australian Journal of Agricultural and Resource Economics, 52, 253-282.

Scarpa, R. & Thiene, M. (2005). Destination choice models for rock climbing in the Northeastern Alps: A latent-class approach based on intensity of preferences. Land Economics, 81, 426-444.
Ščasný, M. & Alberini, A. (2012). Valuation of mortality risk attributable to climate change: Investigating the effect of survey administration modes on a VSL. International Journal of Environmental Research and Public Health, 9, 4760-4781.
Schultz, E.T., Johnston, R. J., Segerson, K. & Besedin, E. Y. (2012). Integrating ecology and economics for restoration: Using ecological indicators in valuation of ecosystem services. Restoration Ecology, 20, 304-310.
Shonkwiler, J. S. & Shaw, W. D. (1997). Shaken, not stirred: A finite mixture approach to analyzing income effects in random utility models. Paper presented at the 1997 Annual Meeting of the American Agricultural Economics Association. August 2-4, Toronto, Ontario.
Small, K. A. & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. Econometrica, 49, 105-130.
Smith, V. K. & Desvousges, W. H. (1986). Measuring water quality benefits. Boston: Kluwer-Nijhoff.
Street, D. J. & Burgess, L. (2007). The construction of optimal stated choice experiments: Theory and methods. Hoboken, NJ: Wiley-Interscience.
Street, D. J., Burgess, L. & Louviere, J. J. (2005). Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. International Journal of Research in Marketing, 22, 459-470.
Sur, D., Cook, J., Chatterjee, S., Deen, J. & Whittington, D. (2007). Increasing the transparency of stated choice studies for policy analysis: Designing experiments to produce raw response graphs. Journal of Policy Analysis and Management, 26, 189-199.
Swait, J. (1994). A structural equation model of latent segmentation and product choice for cross-sectional, revealed preference choice data. Journal of Retailing and Consumer Services, 1, 77-89.
Swait, J. & Adamowicz, W. (2001a). Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organizational Behavior and Human Decision Processes, 86, 141-167.
Swait, J. & Adamowicz, W. (2001b). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. Journal of Consumer Research, 28, 135-148.
Swait, J. & Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research, 30, 305-314.
Thurstone, L. L. (1927). A law of comparative judgment. Psychology Review, 34, 273-286.
Torgerson, W. S. (1958). Theory and methods of scaling. New York: Wiley.
Train, K. (2003). Discrete choice methods with simulation. New York: Cambridge University Press.
Train, K. E. (1998). Recreation demand models with taste differences over people. Land Economics, 74, 230-239.
von Haefen, R. H. & Phaneuf, D. J. (2008). Identifying demand parameters in the presence of unobservables: A combined revealed and stated preference. Journal of Environmental Economics and Management, 56, 19-32.
Vossler, C. A. 2016. Chamberlin meets Ciriacy-Wantrup: Using insights from experimental economics to inform stated preference research. Canadian Journal of Agricultural Economics, 64, 33-48.
Vossler, C. A., Doyon, M. & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. American Economic Journal: Microeconomics, 4 (4), 145-171.
Yellot, J. I., Jr. (1977). The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution. Journal of Mathematical Psychology, 15, 109-144.
Zhang, J. & Adamowicz, W. L. (2011). Unravelling the choice format effect: A context-dependent random utility model. Land Economics, 87, 730-743.

Chapter 6
Travel Cost Models

George R. Parsons

Abstract This chapter provides an introduction to Travel Cost Models used to estimate recreation demand and value recreational uses of the environment such as fishing, rock climbing, hunting, boating, etc. It includes a brief history, covers single-site and random-utility-based models, and discusses current issues and topics. The chapter is laid out in a step-by-step primer fashion. It is suitable for first-timers learning about travel cost modeling as well as seasoned analysts looking for a refresher on current approaches. The chapter includes an application of the random-utility-based model to beach use on the east coast of the USA along with measures of welfare loss for beach closures and changes in beach width.

Keywords Travel cost model · Recreation demand · Valuation · Welfare analysis · Random utility · Per-trip loss

Economists have been concerned with measuring the economic value of recreational uses of the environment for more than 50 years. This has been motivated largely by benefit-cost analyses of environmental regulations and litigation where laws call for valuation in circumstances where harm has been done to the environment. The travel cost model (TCM) is the revealed preference method used in this context.

The basic insight underlying the TCM is that an individual's "price" for recreation at a site, such as hiking in a park or fishing at a lake, is his or her trip cost of reaching the site. Viewed in this way, individuals reveal their willingness to pay for recreational uses of the environment in the number of trips they make and the sites they choose to visit. The TCM is used to value site access and quality change at recreation sites. An example of site access is the closure of a beach due to an oil spill, where access to the site is lost. An example of a quality change is an improvement in water quality on a lake used for recreation.

G.R. Parsons (✉)
University of Delaware, Newark, DE, USA
e-mail: gparsons@udel.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_6

It is common to classify TCMs into two groups: single-site models and random utility maximization models.[1] The earliest TCMs are single-site models that work much like a demand function for any consumer good. Trip cost is treated as the price of the good, and the number of trips taken over a season is treated as the quantity demanded. The simple observation that the closer one lives to a site (lower price), the more trips one takes (higher quantity demanded) is, in effect, viewed as a downward-sloping demand function. People reveal their access value of a site in the number of trips taken.

Random utility maximization (RUM) models consider an individual's choice of visiting one recreation site from among many possible sites on a given choice occasion. The site a person visits is assumed to be a function of the attributes of the sites (size, quality, access, etc.) and the trip cost of reaching the site. Individuals reveal their relative values of site attributes in the sites they choose.

This chapter begins with a brief history of the TCM, followed by detailed presentations of the single-site and RUM models. Next is a step-by-step discussion of estimating a RUM model, which is followed by a short application. The chapter closes with a discussion of some other considerations and a conclusion. There are many good review articles and edited books on travel cost models worth reading along with this chapter (Haab and McConnell 2002, Chapters 6-8; Phaneuf and Smith 2005; Freeman et al. 2014, Chapter 9; Herriges and Kling 1999, 2008; Hanley et al. 2003a, b; Whitehead et al. 2011; Parsons 2013).

6.1 A Brief History of the Travel Cost Model

The earliest travel cost models, dating from the late 1950s and into the 1960s, used "zonal" data and followed a method proposed by Hotelling (1949). Geographic zones were defined around a single recreation site. The zones might be concentric or otherwise spatially delineated but varied in their distance from the site. Visitation rates to the site—observed visits from the zone divided by population in the zone—were computed for each zone. The visitation rate would naturally decline with distance from the site. Travel cost would be computed for each zone based on travel to the center of the zone or the zone's major population center. Then, regressing visitation rate on travel cost, one could predict how that rate would fall with travel cost and, in turn, infer an aggregate demand function. The zonal model ruled for the first decade of the TCM's existence (Clawson and Knetsch 1969; Trice and Wood 1958).

Although the zonal method sees some use in current practice (Moeltner 2003), in the 1970s most modeling transitioned from aggregate zonal data to individual-based data as the unit of observation (Brown and Nawas 1973). The transition

[1]A third group is the Kuhn-Tucker model, which combines features from the single-site and RUM models. It is not covered in this chapter. It is used less frequently and is more advanced than needed in this textbook. Phaneuf and Siderelis (2003) provide an excellent primer-like introduction to the Kuhn-Tucker model.

transformed the method as the tie to consumer theory became clearer, and the behavioral detail that could be studied increased dramatically. Around the same time, researchers became interested in applications to multiple sites in a demand system (Burt and Brewer 1971; Cicchetti et al. 1976). Most of applied microeconomics at the time was being done in a setting with a system of demand equations, and it seemed critical that the TCM should account for substitute sites. In a rather influential article, Hof and King (1982) showed that a full-blown demand system was not needed if the policy interests were narrowly defined on one site. Analysts needed to account for substitute trip costs in the demand model for a single site, but did not need to estimate demand over many sites. This insight is used to this day in single-site models.

Most research in the 1960s and '70s was motivated by valuing site access or per-trip values that could be used in preservation versus development decisions or large water resource projects. In the late 1970s and early '80s interest turned to valuing quality changes at recreation sites such as improvements in water quality in the Chesapeake Bay. Research moved on several fronts to meet this demand. Some analysts attempted pooling or stacking single-site models with variation in quality across sites to estimate shifts in demand (Smith and Desvousges 1985). Some considered a hedonic variant of the TCM that regressed trip cost on-site characteristics over a set of sites (Brown and Mendelsohn 1984). Some explored demand systems using share equations that included quality attributes (Morey 1981). Others adopted the random utility theory in a TCM framework (Bockstael et al. 1984; McFadden 2001).

The RUM model was the popular choice among these in the 1980s and into the '90s, with seminal works by Bockstael et al. (1984, 1987) on beach use and Carson et al. (1987, 2009) on recreational fishing in Alaska. These large-scale efforts paved the way for much of the RUM research that followed. The RUM model proved to be enormously flexible and had an underlying "story of site choice" that appealed to researchers and decision-makers alike. Most importantly, it allowed researchers to value changes in quality or site closures at many sites (in some cases, for a very large number of sites).

In the late 1980s and the '90s, researchers began exploring a wide variety of issues in the context of the RUM model: choice set formation (Feather 1994; Parsons and Kealy 1992), linking trip frequency with site choice (Bockstael et al. 1987; Hausman et al. 1995; Morey et al. 1991), nested logit (Hauber and Parsons 2000), integrating the model with stated-preference data (Adamowicz et al. 1997), considering intertemporal decisions (Adamowicz 1994), and more. The literature also swelled in application to a wide variety of recreation types: fishing, swimming, boating, climbing, beach use, hiking, hunting, skiing, and more.

At the same time the RUM model was taking off, the traditional single-site model was undergoing a transformation of its own. The earliest versions were confined to the world of ordinary least squares, treating trips as a continuous variable. Beginning in the late 1980s and into the '90s, researchers began using limited dependent variable and count data models (Hellerstein and Mendelsohn 1993; Hellerstein 1991, 1992; Shonkwiler and Shaw 1996; Haab and McConnell

1996; Shaw 1988). These models accounted for several subtle aspects of seasonal data and of behavior: the integer nature of trip data, the tendency of trip data to have a frequency distribution skewed toward lower trip counts, the truncation of data due to on-site data collection, and the treatment of participation in recreation. Count data models and on-site data collection became the standard. Analysts also began exploring systems of demand equations using count data models (Englin et al. 1998).

In the last two decades, RUM and single-site models have undergone considerable refinement. In the case of the RUM model, researchers started using simulated probability models and mixed logit (Train 1998). This allowed one to break the stranglehold of the independence of irrelevant alternatives and to explore unobserved heterogeneity. While the nested models introduced in the 1980s allowed researchers to break away from the independence of irrelevant alternatives, at best they only accounted for fairly coarse patterns of substitution. Analysts also began using latent class models, which fall into the mixed logit grouping, to account for heterogeneity across samples. Other developments followed: models for handling on-site sampling (Hindsley et al. 2011), the use of alternative specific constants to control for unobserved variables (Murdock 2006), instrumental variables to deal with issues of congestion (Timmins and Murdock 2007), explicit integration with bio-economic models (Massey et al. 2006), and consideration of modeling in willingness-to-pay space (Scarpa et al. 2008).

On the single-site side, researchers began using the model in conjunction with stated-preference data to measure changes in quality at recreation sites that users had not experienced but might in the event of some policy action. This became the major use for the single-site model in the 2000s and along with it came econometric issues associated with panel data (Englin and Cameron 1996; Whitehead et al. 2000; Landry and Liu 2011). The RUM model also saw increased use combined with stated-preference data but not to the same extent as the single-site model—at least in the context of the TCM (Jakus et al. 2011).

Finally, beginning in the late 1990s and into the early 2000s, Kuhn-Tucker models came on the scene (Phaneuf et al. 2000). These were motivated by a desire to integrate seasonal and site choice models into a unified utility theoretic framework. As noted above, this chapter will not discuss Kuhn–Tucker models.

6.2 Theory and Estimation

This section covers the theory and estimation of single-site and RUM travel cost models. The traditional single-site model is best used when policy issues pertain to one site. When policy issues pertain to several sites and substitution across sites is likely, the RUM model is preferred, and it is now the most commonly used TCM.

6.2.1 Single-Site Model

6.2.1.1 Theory

The single-site model considers demand for a recreation site over an entire season (Bin et al. 2005; Sohngen 2000). Trips to a site are treated as the "quantity demanded" and the trip cost is treated as the "price." This gives a conventional demand function representation as

$$x_n = f(p_n, R_n, Z_n, y_n), \tag{6.1}$$

where x_n is the number of trips taken by person n to the site during the season, p_n is the price or trip cost for the individual to reach the site (travel and time cost), R_n is a vector of trip costs to substitute sites, Z_n is a vector of individual characteristics believed to influence the number of trips taken in a season (e.g., age, family size), and y_n is income. If the number of trips taken declines with the distance one is located from the recreation site, Eq. (6.1) is a downward-sloping demand function. The basic form can be derived from utility theory, wherein one typically uses separate budget and time constraints to show explicitly the opportunity cost of time as a part of trip cost (Haab and McConnell 2002, p. 142).

Using Eq. (6.1), consumer surplus for access to the site is the integral

$$cs_n^{\text{site}} = \int_{p_n^0}^{p_n^*} f(p_n, R_n, Z_n, y_n)\,\mathrm{d}p_n, \tag{6.2}$$

where p_n^0 is current trip cost to the site and p_n^* is a choke price (the trip cost where demand for trips goes to zero). Figure 6.1 shows the surplus (Area A) graphically. If

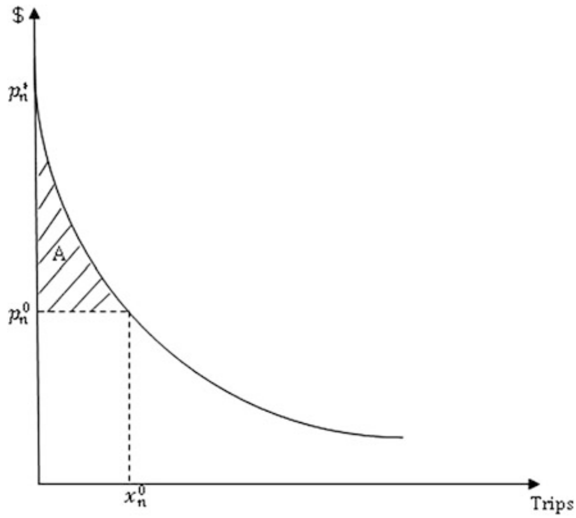


Fig. 6.1 Access value in a single-site model

the site were lost, for example, cs_n^{site} would be the loss in welfare to person n. This is often called lost "access value". Analysts sometimes report mean per-trip values as cs_n^{site}/x_n. An exact Hicksian measure of surplus may be derived as well (see Chap. 2, Sects. 2.1 and 2.2). The log-linear form for Eq. (6.1) is popular for the single-site model:

$$\ln(x_n) = \alpha p_n + \beta_R R_n + \beta_z Z_n + \beta_y y_n. \tag{6.3}$$

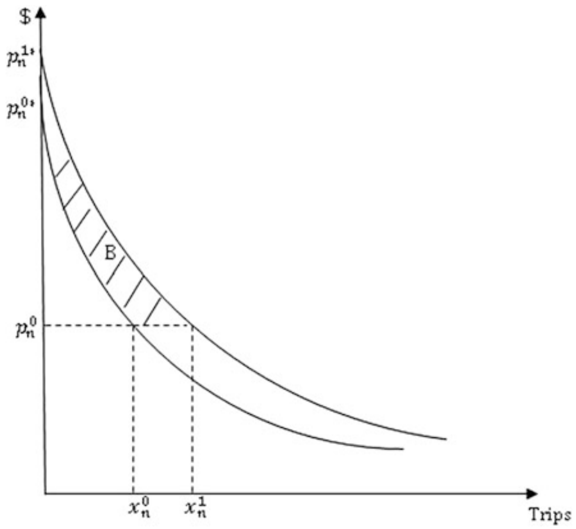
In this case, seasonal access value in Eq. (6.2) has the convenient form $x_n/-\alpha$, and per-trip values (dividing the seasonal value by x_n) are simply $1/-\alpha$, which is constant across individuals.

The model is also used to value changes in site quality (improved fishing or more hiking trails). For this, one requires behavioral data on recreational use under different conditions of quality. The most commonly used is stated-preference data wherein the researcher asks respondents in a TCM survey how their number of trips would change under different conditions of quality. Then, the analyst considers an individual's demand function with and without a change in quality. The "with and without" difference in consumer surpluses measures the value of the quality change

$$cs_n^{\text{Quality}} = \int_{p_n^1}^{p_n^{1*}} f^1\left(p_n, R_n, Z_n, y_n, q^1\right) \mathrm{d}p_n - \int_{p_n^0}^{p_n^{0*}} f^0\left(p_n, R_n, Z_n, y_n, q^0\right) \mathrm{d}p_n, \tag{6.4}$$

where, for example, q^0 is initial water quality at a recreation site and q^1 is water quality after some improvement. This result uses the theory of "weak complementarity" presented by Flores in Chap. 2, Sect. 2.2.2, which holds that one must

Fig. 6.2 Quality-change value in a single-site model



take a trip to the site to enjoy the benefits of a quality improvement there. Figure 6.2 shows the gain in surplus (Area B) graphically. It is the "added" area due to the shift in demand. In a log-linear demand model in the simplest representation of quality, q enters the right side of Eq. (6.3) as $+\beta_q q$ and the seasonal and per-trip values for a quality change are $\left(x_n^1 - x_n^0\right)/-\alpha$ and $\left(x_n^1 - x_n^0\right)/-\alpha x_n^0$, where x_n^1 is the quantity demanded when $q = q^1$, and x_n^0 is the quantity demanded when $q = q^0$. The same analysis can be done for declines in quality where the demand function will shift to the left instead of to the right, and the area between the demand curves will be interpreted as a loss instead of a gain.

This approach, in effect, combines revealed preference (trips actually taken) with stated-preference (trips a person "states" he or she will take if quality were to change). For more on stated-preference analysis, see Chaps. 4 and 5. For more on combining revealed and stated-preference analysis, see Whitehead et al. (2011). For a recent example see Parsons et al. (2013).

6.2.1.2 Estimation

In estimation, the analyst gathers cross-sectional data on a sample of individuals for a given season: number of trips taken to the site, trip cost, and other relevant demand shifters. Then, via the spatial variation in trip cost (people living at different distances from the site and so having different prices for trips), an equation like (6.3) is estimated. If one is measuring a quality change using contingent-behavior data, estimation also accounts for demand shifts under the different quality conditions—essentially estimating the effect of a variable that shifts the demand curve. After the model is estimated, access values and quality change values can be computed using Eqs. (6.2) and (6.4).

Choice of functional form for estimation has been the subject of inquiry since the TCM methodology was first developed. Earlier functional forms were continuous: linear, log-linear, log-log, etc. Most modern forms are count data models—Poisson, negative binomial, zero-inflated, hurdle, etc.—and use a log-linear expected value function for demand. Count data models are designed for analyses with a dependent variable measured as a nonnegative integer and are quite versatile for handling truncation, large numbers of zero trips in the data, and preference heterogeneity. This has made them popular for single-site demand function estimation (Hellerstein 1991; Creel and Loomis 1990; Hellerstein and Mendelsohn 1993).

A Poisson model is the simplest count data model. An individual's probability of making x_n trips to a site in a given season in the Poisson model is

$$\mathrm{pr}(x_n) = \frac{\exp(-\lambda_n) \cdot \lambda_n^{x_n}}{x_n!}, \tag{6.5a}$$

$$\ln(\lambda_n) = \alpha p_n + \beta_R R_n + \beta_Z Z_n + \beta_y y_n, \tag{6.5b}$$

where λ_n is the expected number of trips taken by person n and $\ln(\lambda_n)$ is a log-linear specification like Eq. (6.3). Essentially, one is estimating the parameters of the demand function in Eq. (6.5b). The Poisson form in Eq. (6.5a) is forcing the outcomes (trips in this case) to be nonnegative integers. The parameters are estimated by maximum likelihood, where each person's probability of taking the number of trips actually taken is used as an entry in the likelihood function. Seasonal consumer surplus (Eq. (6.2) and Area A in Fig. 6.1) in the Poisson form is $\widehat{S} = \hat{\lambda}_n / -\hat{\alpha}$, where the hatted values denote estimates, and again λ_n is the expected number of trips by person n. Per-trip consumer surplus is $\hat{s}_n = \widehat{S}/\hat{\lambda}_n = 1/-\hat{\alpha}$.

An undesirable feature of the Poisson model is that the mean and variance of x_n are constrained to be equal. After testing, if the data fail to support this assumption, a negative binomial model is commonly used to relax this constraint. Testing often shows that the mean and variance are not equal (Greene 2007, p. 911; Haab and McConnell 2002, p. 169).

6.2.1.3 On-Site and Off-Site Samples

One of the more important decisions to make when estimating a single-site model is whether to gather data on-site or off-site. Often, the number of visitors to a particular recreation site is a small fraction of the general population and there is no identifier for people who visit the park. If so, sampling the general population may require a large number of contacts to form a reasonable sample size of current and potential visitors. On-site data has the advantage that every individual intercepted has taken at least one recreation trip. In this way gathering data on-site is usually cheaper than collecting it off-site.

However, there are at least two disadvantages of on-site data: endogenous stratification and truncation. Because individuals taking more trips over a season are more likely to be drawn for inclusion in the sample, there is oversampling in direct proportion to the number of trips one takes over a season (i.e., a person taking two trips is twice as likely to be sampled as a person taking one trip). Estimation that fails to account for this effect, which is known as endogenous stratification, will give biased parameter estimates for a model intended to represent the general population. Also, with on-site sampling, the analyst never observes individuals taking zero trips in the season, so there is no direct observation at the choke price on the demand function. The dependent variable is said to be truncated, and this also needs to be accounted for when using the standard model. Both endogenous stratification and truncation are easily corrected econometrically. One can show that using $x_n - 1$ instead of x_n in estimation in Eq. (6.5a) corrects for both effects (Shaw 1988; Haab and McConnell 2002, p. 174). Correcting for on-site sampling in the negative binomial model does not have the convenient "$x_n - 1$" result, but corrected forms do exist (Englin and Shonkwiler 1995; Martinez-Espineira and Amoako-Tuffour 2008; Edwards et al. 2011).

Off-site sampling has the advantage that nonparticipants are observed. This presumably makes for more accurate estimation of the choke price and, hence, estimation of consumer surplus. In some models, nonparticipants get special treatment, wherein the analyst estimates a two-part model. First, the decision to take a recreation trip or not (the participation decision), and second, how many trips to take (the frequency decision). These models are known as hurdle or zero-inflated Poisson models. The models, more or less, use a simple bivariate choice model for the participation decision and a Poisson model of one form or another for the frequency decision. If one believes that participation and frequency are governed by different decision processes, these models are beneficial. A zero-inflated Poisson model applied in our context has the form

$$\mathrm{pr}(x_n = 0) = \Phi(Z_n, y_n) + (1 - \Phi(Z_n, y_n)) \exp(-\lambda_n), \tag{6.6a}$$

$$\mathrm{pr}(x_n | x_n > 0) = (1 - \Phi(Z_n, y_n)) \frac{\exp(-\lambda_n)\lambda_n^{x_n}}{x_n!}. \tag{6.6b}$$

The first Eq. (6.6a) is the probability of observing a person take zero trips in the season, and the second Eq. (6.6b) is the probability of observing a person take one or more trips in the season where λ_n is the same as shown in Eqs. (6.5a and 6.5b) and Φ_n is a simple bivariate-logit model. The first term in the Eq. (6.6a) models the participation decision—the probability of a person not being someone who engages in the type of recreation under study at all. The second term in Eq. (6.6a) captures those who engage in the type of recreation under study but happen to not make a trip in the current season—the probability of being a participant but taking no trips. Equation (6.6b) is the frequency decision for those taking at least one trip—the probability of taking trips in the season given that the person is a participant. Again, estimation is by maximum likelihood, wherein probabilities from Eqs. (6.6a and 6.6b) are loaded for each observation according to actual choices. Seasonal and per-trip consumer surplus in a zero-inflated model have the forms $\left(1 - \widehat{\Phi}_n\right)\lambda_n / -\hat{\alpha}$ and $1/-\hat{\alpha}$, where the "weighting" term, $(1 - \widehat{\Phi}_n)$, accounts for participation (Haab and McConnell 1996; Gurmu and Trivedi 1996).[2]

6.2.1.4 Valuing a Quality Change

Now consider the application of the single-site model for valuing a quality change at a site. In principle, the analyst estimates two demand models: one using reported trips during the season and a second using reported trips based on some hypothetical change in quality at the site. Then, for any given respondent, the analyst differences the consumer surplus or access value estimated for the two models to

[2]For an interesting discussion of using zero inflated Poisson models see Paul Allison's commentary at www.statisticalhorizons.com/zero-inflated-models.

arrive at the value of the quality change—the area between the two demand curves (Fig. 6.2). It is common to estimate the two models simultaneously, and in some cases, to constrain coefficients to be the same across the models. At one extreme, the analyst might constrain all coefficients but the constants to be equal, allowing only for a shift in demand. At the other extreme, the analyst might let all coefficients vary. The latter allows the effect of the quality to vary by individual characteristics. So, for example, the effect of water quality on demand might vary by income and age. Sometimes the quality change is presented in terms of future trips. In this case, the analyst obtains data on trips in the current season, expected trips in the next season without a quality change, and expected trips in the next season with a quality change. Now there are three demand equations to estimate.

Most analysts attempt to account for the panel nature of the contingent-behavior data. That is, the error terms in the revealed preference demand equation and the stated-preference demand equation are likely to be correlated for each respondent. The random effects Poisson model (aka Multivariate Poisson Gamma) is a common form that allows for this correlation (Hanley et al. 2003a, b; Whitehead et al. 2008; Morgan and Huth 2011; Parsons et al. 2013). The expected trip function in Eq. (6.5b) now takes the form

$$\ln\left(\lambda_{nj}\right) = \alpha_j p_n + \beta_{Rj} R_n + \beta_{zj} Z_n + \beta_{yj} y_n + \beta_{qj} q_j + \varepsilon_{nj}, \tag{6.7}$$

where $j = 0$ denotes current condition (without quality change), and $j = 1$ denotes the contingent condition (with quality change). The random effects Poisson sets $\varepsilon_{nj} = \varepsilon_n$ so each individual's error term is perfectly correlated. This is the "random effect" shared across each choice made by one person. If $\exp(\varepsilon_n)$ is distributed as a normalized Gamma, the result is a random effects Poisson model. The "$x_n - 1$" trick mentioned above can be used with on-site samples in random effects Poisson models, which is a nice simplifying feature. Its major drawback is that the manner in which correlation is captured is rather restrictive. There are a number of applications using more advanced forms (Landry and Liu 2009, 2011; Egan and Herriges 2006; Awondo et al. 2011).

6.2.2 Random Utility Maximization (RUM) Model

6.2.2.1 Theory

The RUM model is a multiple-site model that explicitly accounts for substitution among sites and easily accommodates access and quality change valuation. The RUM model tells a more complete behavioral story than the single-site model; it accounts for site choice and frequency of trips over a season. It also accommodates quality change valuation without the addition of stated-preference data.[3] Flores (Chap. 2, Sect. 2.3.3) covers the theory of valuation in a RUM model and is a good cross-reference for this section.

The time frame for a RUM model is a single choice occasion, usually a day, in which an individual makes one recreation trip. The individual is assumed to face a set of possible sites for a trip. The sites might be rivers, parks, beaches, etc. Each site $i(i = 1, 2, \ldots, I)$ is assumed to give an individual some utility U_{in} on a given choice occasion. *Site utility* is usually assumed to have a linear form:

$$U_{in} = \alpha p_{in} + \beta_q Q_i + \varepsilon_{in}, \tag{6.8}$$

where p_{in} is trip cost and Q_i is a vector of site attributes (qualities). As above, trip cost includes travel and time cost and is treated as a person's "price" of recreation. Site attributes include qualities such as natural amenities, water quality, size, access, and so forth. One expects site utility to rise with desirable attributes, such as higher water quality, and decline with undesirable attributes, such as lower catch rates of fish. The error term, ε_{in}, captures site attributes and individual characteristics that influence site choice but are unobserved by the analyst.

Because preferences may vary across individuals, analysts sometimes introduce heterogeneity into the model. This is done with interaction terms:

$$U_{in} = \alpha p_{in} + \beta_q Q_i + \beta_{qz} \widetilde{Q}_i \widetilde{Z}_n + \varepsilon_{in}, \tag{6.9}$$

where $\widetilde{Q}_i$ is a vector of site attributes and $\widetilde{Z}_n$ is a vector of individual characteristics, such as age, boat ownership, etc. The interaction terms allow site characteristics in $\widetilde{Q}_i$, which may share many of the same variables as Q_i, to affect people differently. For example, a boat ramp as a site attribute in $\widetilde{Q}_i$ might be interacted with boat ownership as an individual characteristic in $\widetilde{Z}_n$ if one believes that boat ramps only matter to those who own boats.

On any given choice occasion then, an individual is assumed to choose the site with the highest site utility, giving *trip utility*:

$$V_n = \max(U_{1n}, U_{2n}, \ldots, U_{In}). \tag{6.10}$$

Trip utility, V_n, is the basis for welfare analysis in the RUM model. It is used to value a loss or a gain from site access (removal or addition of a site) and changes in site quality. For example, consider an oil spill that closes sites 1 and 2. Trip utility with the closures becomes

$$V_n^{\text{closure}} = \max(U_{3n}, U_{4n}, \ldots, U_{In}), \tag{6.11}$$

[3]The single-site model can also be used to value quality change without stated-preference data by "pooling" or "stacking" many separate single-site models (Smith and Desvousges 1985). However, this approach does not account for substitution among sites and has largely fallen out of favor.

where sites 1 and 2 have been dropped from the choice set. Trip utility then declines from V_n to V_n^{closure}. A similar expression can be generated for a change in site quality at one or more sites. Suppose the water quality at sites 2 and 3 is improved through some regulation. If so, trip utility for person n becomes

$$V_n^{\text{clean}} = \max\left(U_{1n}, U_{2n}^*, U_{3n}^*, U_{4n}, \ldots, U_{In}\right), \tag{6.12}$$

where U_{2n}^* and U_{3n}^* denote the now higher utility due to the improved quality. In this case, trip utility increases from V_n to V_n^{clean}. In both cases, the change in utility in the simple linear model is monetized by dividing the change by the negative of the coefficient on trip cost $(-\alpha)$, which is our marginal utility of income.[4] This gives the following compensating variation (also equivalent variation) measures for changes in trip utility:

$$\begin{aligned} w_n^{\text{closure}} &= \left(V_n^{\text{closure}} - V_n\right)/-\alpha, \\ w_n^{\text{clean}} &= \left(V_n^{\text{clean}} - V_n\right)/-\alpha. \end{aligned} \tag{6.13}$$

These are changes in welfare on a per-trip, per-person basis. Notice the behavior underlying the welfare measures in Eqs. (6.9, 6.10, 6.11, 6.12, and 6.13). If an individual's preferred site (i.e., site yielding maximum utility) changes as the choice set changes due to closure or quality change, the change in trip utility is just the difference in the utility at the preferred site without the change and the preferred site with the change. If the individual's preferred site is the same (with and without the closure or quality change), the change in trip utility is the difference in that site utility with and without the change. In this way, substitution across sites drives values in a RUM model (for examples, see Phaneuf, 2002, Murray et al., 2001, and Parsons et al., 2009).

6.2.2.2 Estimation

The error terms, ε_{in}, on each site utility are unknown to researchers. For this reason, the choice is treated as the outcome of a stochastic process in estimation. By assuming an explicit distribution for the error terms in Eq. (6.9), we can express each person's choice as a probability of visiting each site in the choice set. The simplest is to assume that the errors terms are independently and identically distributed Type 1 extreme value random variables. This results in a multinomial logit specification for the choice probabilities (Greene 2007, Chapter 23). Each person's probability of choosing any site k from the set of sites is

[4]The coefficient α in Eqs. (6.8) and (6.9) is a measure of the marginal utility of income because it describes how site utility changes with a decrease in income (less money to spend on other things) if a trip is taken. Because trip cost "takes away from income," α is the marginal effect of taking away income ($\alpha < 0$), and $-\alpha$ is a measure of adding to income or the marginal utility of income ($-\alpha > 0$).

$$\mathrm{pr}_n(k) = \frac{\exp\left(\alpha p_{kn} + \beta_q Q_k + \beta_{qz} \widetilde{Q}_k \widetilde{Z}_n\right)}{\sum_{i \in I} \exp\left(\alpha p_{in} + \beta_q Q_i + \beta_{qz} \widetilde{Q}_i \widetilde{Z}_n\right)}. \tag{6.14}$$

The parameters are then estimated by maximum likelihood using data on actual site choices. Because researchers precede as though choices are the outcomes of a stochastic process, trip utility in Eqs. (6.9) through (6.13) is also stochastic. For this reason, expected trip utility is used as an estimate of V_n in empirical work. It can be shown that each individual's expected trip utility in a multinomial logit model is

$$\begin{aligned} E(V_n) &= E\{\max(U_{1n}, U_{2n}, \ldots, U_{In})\} \\ &= \ln\left\{\sum_{i=1}^{I} \exp\left(\alpha p_{\mathrm{in}} + \beta_q Q_i + \beta_{qz} \widetilde{Q}_i \widetilde{Z}_n\right)\right\} + C, \end{aligned} \tag{6.15}$$

where C is some unknown additive constant (Small and Rosen 1981). $E(V_n)$ is often referred to as the "log sum" and again is the empirical form of V_n used in welfare analysis. The steps in such an analysis are straightforward. One estimates the parameters of site utility, uses the parameters to construct expected trip utilities with and without some resource change using Eq. (6.15), computes per-trip changes in utility per person, substituting $E(V_n)$ for V_n with and without the change in Eq. (6.13), and finally, using the estimate of α to monetize the change in expected utility (C drops out when one differences the equations).[5]

One of the major drawbacks of the multinomial logit model is the restrictive way in which substitution occurs. Since site substitution is the pathway through which welfare effects are captured, it is important to handle it in as realistic a way as possible. The multinomial logit model assumes that the closure or decline in quality at one or more sites leads to a proportional increase in the visitation to all other sites—their shares remain in fixed proportion. This property, known as the independence of irrelevant alternatives, is usually unrealistic (Train 2009, p. 45). For this reason, economists have turned to alternative forms that allow for more realistic patterns of substitution. This is achieved, at least in a stochastic sense, by allowing for correlated error terms across the sites in Eq. (6.9). There are two common methods that allow for such correlation: nested and mixed logit specifications.

The nested logit model has been used since the introduction of travel cost RUM models (Bockstael et al. 1987; Morey 1999; Grijalva et al. 2002; Cutter et al. 2007). Nested models place sites that share unobserved characteristics in a common "nest" under the assumption that they serve as better substitutes for one another than sites outside the nest. This model includes a new parameter for each nest that captures the degree of substitution among the sites within that nest. Researchers often nest

[5]In some cases, researchers will consider site utilities that are nonlinear in trip cost, allowing for nonconstant marginal utility of income and empirical forms of Eq. (6.15) that are not closed-form. See Herriges and Kling (1999) for a discussion and example.

sites by proximity (e.g., grouping sites in the same region together), resource type (e.g., grouping lakes and rivers in separate nests), and purpose (e.g., grouping trout and bass fishing trips separately). The latter groupings are an example of expanding the choice model to include not only site choice but also recreation activity choice, which is common in travel cost RUM models and may enhance the policy relevance of value estimates.

The mixed logit model (or random parameters logit) is a more flexible approach for enriching the patterns of substitution (Train 1998; Jeon et al. 2011). It not only allows for more and overlapping substitution structures, it can also be easily configured to mimic what a nested logit model does. The mixed logit probability is an integral over a basic logit probability

$$\mathrm{pr}_n^*(k) = \int Lgt_k(\alpha, \beta) g(\beta|\mu, \varphi)\, d\beta, \tag{6.16}$$

where Lgt_k is a logit probability from Eq. (6.14) and $g(\beta|\mu, \varphi)$ is a mixing distribution (may be normal, triangular, or some other distribution) with mean μ and standard deviation φ. Now, the analyst seeks to estimate $\alpha, \mu,$ and φ. That is, instead of estimating a simple scalar for each element in β, he or she seeks to estimate a distribution for each parameter.[6] In this way, it is a “random parameter” model. The integral in Eq. (6.16) is estimated by simulated probability (see Train, 2009, Chapter 6).

The mixed logit model allows sites to share correlation through the term φ, which allows for more complex substitution across sites. For example, suppose two sites in a choice set share the attribute “park” (park = 1 if the site is a state park and = 0 if it is not). The element in the parameter vector corresponding to park picks up the degree of correlation between the two sites because an individual’s utilities at the two sites move in concert with the outcome for that parameter. This introduces substitution in much the same way as a nested model but can do so with more flexibility across many different and overlapping attributes. Expected trip utility in the mixed logit model takes the same form as in the standard logit model (Eq. (6.15)). However, because the parameters are now random, $E(V_n)$ is treated as an average estimated by randomly drawing from the estimated parameter distributions (Parsons and Massey 2003).

An alternative interpretation of the mixed logit model is that its random parameters pick up unobserved heterogeneity across the sample (Provencher and Moore 2006; Thiene and Scarpa 2009; Breffle and Morey 2000). The variation in this case is capturing variety in preference. Unlike the observed heterogeneity discussed in the interactions in Eq. (6.9), this heterogeneity is captured in an error term distribution and is therefore described as being unobserved.

---

[6]Following convention, I have specified α, the coefficient on trip cost, as fixed. Because α is used in the denominator of Eq. (6.13) for valuation, values tend to be extremely sensitive to variation created by mixing. This is a practical fix and an area where more research is needed.

6.2.2.3 On-Site and Off-Site Samples

Like the single-site model, the choice of using on-site versus off-site data is a major decision that affects the econometrics used to estimate the model. Sometimes on-site sampling is desirable. For example, when studying a set of sites with low visitation relative to other sites in the choice set, one might heavily sample there to obtain observational data on the sites of interest for policy. Or, to increase sample size, one might sample more popular sites or easily accessible sites more heavily than other sites. The difficulty with on-site sampling in the RUM context is that the relative frequency with which people choose different sites is the behavior the researcher is modeling. If the sampling strategy across sites is not random, the data is mixing behavioral choice with researcher-sampling choice, and parameter estimates will be biased. The researcher, in effect, thinks the more heavily sampled sites are more popular than they actually are. The sampling choice is contaminating the observation data. This is referred to as choice-based sampling. In addition, the same issues discussed for the single-site model with regard to on-site sampling apply to RUM models: oversampling of frequent users and having participant-only data. To estimate parameters without bias using on-site data, site weights are needed. The weights are based on the actual distribution of trips by the population of users being studied and are used to adjust the count of trips to the different sites to reflect their true occurrence (Train 1986, pp. 48-49; Ben-Akiva and Lerman 1985; Laitila 1999; Greene 2007, pp. 793-794). There are a number of recent applications of travel cost models using on-site sampling (Grijalva et al. 2002; Moeltner and Shonkwiler 2005; Hindsley et al. 2011; Riera et al. 2011). However, most applications use general population sampling where correction in estimation is not needed.

6.2.2.4 Seasonal Forms

At their core, site choice models are occasion-based, centering on a single trip. Yet, analysts are often interested in the seasonal implications of a policy change or in resource changes that may engender changes in the number of trips taken (e.g., fewer fishing trips if catch rates decline). Two methods are used to modify the basic choice model to make it seasonal and to incorporate the possibility of adjusting the number of trips taken over a season: a repeated discrete choice model with a no-trip alternative and a linked seasonal demand model. These are often called trip frequency models.

The repeated choice model adds a no-trip utility to the individual's choice (Morey et al. 1991; Egan et al. 2009). This takes the form

$$U_{0n} = \delta_0 + \delta_z Z_n + \varepsilon_{0n}, \tag{6.17}$$

where Z_n is again a vector of individual characteristics believed to influence whether or not a person takes a trip on a given choice occasion. This might include age, family composition, years engaged in recreation, and so forth. Each person now has an expanded choice on each choice occasion: visiting one of the I sites or taking no trip. The model is made seasonal by simply repeating it for every choice occasion in the season, where the choice probabilities now include no trip as one of the alternatives. Instead of trip utility, one now uses *choice occasion utility* $V_n^{\text{co}} = \max(U_{0n}, U_{1n}, U_{2n}, \ldots, U_{In})$, where no trip is added as a possible choice, and the log-sum in Eq. (6.15) becomes an expected choice occasion utility:

$$\begin{aligned} E(V_n)^{\text{co}} &= E\{\max(U_{0n}, U_{1n}, \ldots, U_{In})\} \\ &= \ln\left\{\exp(\delta_0 + \delta_z Z_n) + \sum_{i=1}^{I} \exp\left(\alpha p_{in} + \beta_q Q_i + \beta_{qz} \widetilde{Q}_i \widetilde{Z}_n\right)\right\} + C_n. \end{aligned} \tag{6.18}$$

Per-person welfare changes are calculated as before (see Eq. (6.13) and discussion following Eq. (6.15)) but become per-choice occasion (instead of per-trip), because it now includes no trip as part of the choice set. Seasonal estimates of welfare change are simply the per-choice occasion values times the number of choice occasions in a season $M \cdot w_n^{\text{co}}$, where w_n^{co} denotes the per-choice occasion value and M is the number of choice occasions.

Usually, the analyst will have data on trips over an entire season without knowing the specific date for each trip. If so, if a person took trips T on M choice occasions, each of the trips would enter the likelihood function as the probability of taking a trip, and each of the no trips would enter as the probability of taking no trip. This expands the data set considerably in estimation. In nested logit, the sites are usually placed in a nest separate from the no-trip choice. In mixed logit, no-trip utility usually includes its own alternative specific constant, as shown in Eq. (6.17) by δ_0 and is treated as a random parameter. It is also desirable to allow for correlation of utilities across choice occasions in repeated models in estimation (Herriges and Phaneuf 2002).

The alternative approach for introducing a seasonal dimension into a RUM model is a pseudo-demand model that links number of trips taken over a season with the expected trip utility from the RUM model (Bockstael et al. 1987; Hausman et al. 1995; Parsons et al. 2009). The linked model has the form

$$T_n = f((E(V_n)/-\alpha), Z_n), \tag{6.19}$$

where T_n is the number of trips taken by a person over the season, $E(V_n)$ is the expected trip utility estimated in a RUM model divided by the estimated coefficient on trip cost $-\alpha$ from the same model, and Z_n is a vector of individual characteristics. $E(V_n)$ is the expected value of a recreation trip for a person on any given choice occasion.

One would expect the number of trips a person takes to increase with the expected value of a trip. In this way, the model can be used to predict how the number of trips over a season changes with changes in site characteristics or the loss of a site in the choice set. For example, the expansion of designated open space at one or more recreation sites would increase predicted $E(V_n)$ from a RUM model, which, in turn, would increase the number of predicted trips in a linked model, thereby picking up seasonal adjustments in the number of trips taken. The linked model is typically estimated using a count data model and is estimated either stepwise or simultaneously with the RUM model. Essentially, one estimates something like a single-site model, replacing price with $E(V_n)/-\alpha$. Poisson and negative binomial forms are common. A seasonal change in welfare in a Poisson or negative binomial forms is $\widehat{\Delta T_n}/\widehat{\gamma}$, where $\widehat{\Delta T_n}$ is the change in trips due to the resource change and $\widehat{\gamma}$ is the parameter estimate on expected trip utility in the linked model (Parsons et al. 1999a; Herriges et al. 1999).

The model is admittedly ad hoc in the sense that it is not built from a consistent utility theoretic framework at the site choice and trip frequency levels. Nevertheless, it has proven to be quite versatile and usually easy to estimate. Also, the repeated choice model can be written in a linked form as $T_n = M \cdot (1 - \mathrm{pr}(\mathrm{no\_trip}))$, where M is the number of choice occasions in a season and $\mathrm{pr}(\mathrm{no\_trip})$ is the probability of taking no trip. In this way, the two models, while ostensibly different, can be seen as simply different functional forms for the seasonal component of a RUM model.

6.3 Steps in Estimating a RUM Model

This section is a stepwise "how-to" presentation of estimating a RUM model[7]; the steps are shown in Table 6.1. Single-purpose day trips are assumed throughout; the issues of multiple-purpose and overnight trips will be discussed later. While laying out the research in steps helps to clarify the tasks facing the analyst, in truth, the steps are often done more in concert than stepwise. The decisions are so intertwined that it is difficult to think of them in isolation.

6.3.1 Step 1. Identify and Define Impacts

The analysis begins by identifying and defining the impacts to be valued. These will take the form of site closures, openings, or quality changes at one or more sites. Identifying the impacts at the outset defines the research problem and gives direction for the upcoming steps. Typically, the analysis is driven by an anticipated policy event or a need to assess damages at some site or sites. Sometimes a model is estimated in the anticipation of use for a variety of policy questions, and in this

[7]The steps in estimating the single-site model are essentially the same. Site definition (Step 3) is obviously only for one site, and site characteristic data (Step 6) are typically not gathered. In instances where several single-site models are being "stacked" in estimation, analysts will often gather site characteristic data to allow for shifts in demand across sites.

Table 6.1 Steps in estimating a RUM model

Step 1	Identify and define impacts
Step 2	Identify and define recreation uses
Step 3	Identify and define sites
Step 4	Identify population of users and develop a sampling strategy
Step 5	Variable selection
Step 6	Gather site characteristic data
Step 7	Design and implement the survey
Step 8	Measure trip cost
Step 9	Estimate model
Step 10	Report study results

case, the range of applications needs to be considered so that the estimated model will have flexibility for future welfare evaluations.

With a site closure or opening, the analyst identifies the sites that are affected (see Step 3 on defining sites). With a quality change the analyst identifies the affected sites and the changes that will be evaluated. These might be changes in water quality, fish catch, number of hiking trails, presence of a fish consumption advisory, beach width, and so on.[8] It is useful to begin thinking early in the analysis about how quality will be measured for each site. For example, will objective measures of quality be used, such as levels of dissolved oxygen for water quality? Or will perceived measures of quality be used, such as a rating based on the reporting from respondents in a survey (Joen et al. 2011)? Also, are there existing measures of the quality attribute or will new measures have to be constructed? How will these quality changes map into the policy being evaluated? For example, if a policy rule is in terms of loadings of pollutants and a perceived measure of quality is used, how will results be translated into terms of actual loadings?

It is also necessary to establish early that there is sufficient variation in quality across sites to measure the effect on-site choice. If quality is more or less uniform across the sites for the characteristic of interest, it is impossible to measure the effect of quality on-site choice. It is also important to consider whether or not the range of variation covers the range of interest for the policy application. For example, observable beach widths might be in the range of 25 to 75 feet with sufficient variation. If, however, the policy under consideration is to widen beaches to 150 feet, the accuracy of the estimates for this application is limited. If there is a lack of variation or simply no observation on the attribute of interest, augmenting the analysis with stated-preference data is an option (Jakus et al. 2011).

[8]Phaneuf (2002), for example, considers a variety of water quality measures, including pH, dissolved oxygen, phosphorous, ammonia, and an index defined by the U.S. Environmental Protection Agency. Lupi et al. (2003) use a catch rate of fish as a measure of quality.

6.3.2 Step 2. Identify and Define Recreation Uses

Next, one defines the recreation uses affected by the impacts identified in Step 1. In some cases, the sites will have a single or dominant recreation use, such as fishing or beach visitation. In other cases, there will be multiple uses, such as fishing, swimming, and boating. Ideally one would include all recreation types and estimate separate models for each. However, policies or damage assessments sometimes call for focusing on a single recreation type. Government agencies, park services, and tourist bureaus often have information and data to help identify major uses and categories of user types.

If recreation types are similar enough, they may be aggregated. For example, if there are many types of boating, one might treat these as a single group. The more similar the recreation types, the less problematic the aggregation; aggregating sailing and motor boating is probably less problematic than aggregating motor boating and swimming. Most studies have some aggregation. Beach use, which can include sunbathing, swimming, surfing, jogging, and even surf fishing, is often treated as a single recreation type (Parsons et al. 2009; Whitehead et al. 2010). Aggregation simplifies data collection and analysis. Again, one must be careful not to bundle recreation types that are too dissimilar. Parsons and Kealy (1992) used four separate groups for a water-based recreation study: fishing, boating, swimming, and viewing. Feather and Hellerstein (1997) used two groups: lake recreation (including trips to wetlands) and river recreation.

Because individuals are sometimes observed engaging in more than one type of recreation on a single visit, a common practice is to identify the primary purpose of the recreation trip and classify the use accordingly. For example, one might ask respondents in a survey to report the number of trips taken "primarily for the purpose of fishing" and then "primarily for the purpose of bird-watching" and so on.

One must also define the season for each use. Hunting and fishing may have a season defined by law. Skiing, rock climbing, and beach use will have a season defined by periods of favorable weather. Other uses, such a viewing and hiking, may be year-round.

6.3.3 Step 3. Identify and Define Sites

In this step, one defines what a recreation site is, determines the universe of sites to include in the analysis, and, finally, defines which sites to include in each person's choice set.

Outdoor recreation sites take many forms: parks, lakes, beaches, wilderness areas, river segments, etc. Sometimes the boundaries are easy to delineate, such as in the case of parks, lakes, or wildlife reserves (Parsons and Kealy 1992; Day 2000). In other cases, the delineation is not so clear. Rivers for fishing or white-water rafting, for example, call for the analyst to define the length of the segments

(Murdock 2006). Sites for hunting or ocean fishing usually require researcher-defined boundaries as well (McConnell and Strand 1994). A difficulty with marine recreational fishing or boating is that sites may actually be areas in the ocean, which are difficult to delineate, and users may cross several sites on any given trip.

Government agencies, park services, and tourist bureaus often have literature and maps that help in defining sites (Parsons et al. 2009). Sometimes government agencies managing natural resources have their own site definitions, such as wildlife management units (Andrews 1996; Adamowicz et al. 1997). These often work for a travel cost model.

Sites are often aggregated. Some examples are sites defined by hydrologic units or watersheds (Phanuef 2002) or large fishing regions (Hausman et al. 1995) or coastal counties (McConnell and Strand 1994). In choosing the number of sites and the degree of aggregation of sites, it is best to err on the side of many sites and narrow site definitions. Modern econometric software packages can handle a large number of sites. The feasibility of randomly drawing sites in estimation to approximate larger choice sets is also an option (Parsons and Kealy 1992; Feather 1994).

If you must aggregate sites, the general rule is to group similar sites together. The sites should be similar in all characteristics, including trip cost. The less similar the sites, the more bias is likely (Parsons and Needelman 1992; Feather 1994). For example, if choice between sites within an aggregated group happens to be where preferences may be captured for water quality differences, the parameter estimates for true water quality values would be lost because they would be masked through aggregation. Another strategy that is sometimes used is mixing aggregated and nonaggregated sites together in such a way that the bias of aggregation is diminished (Lupi and Feather 1998; Parsons et al. 2000). Also, see Haener et al. (2004) for a more recent treatment of aggregation bias issues in a RUM model.

Once a strategy is developed for defining sites and their boundaries, the next step is to define the universe of sites to include in the analysis. In principle, one wants to include all the sites with impacts identified in Step 1, plus all other sites that may serve as substitutes for these sites for people in the sample identified in the next step. This set is always approximated in some way. Again, political boundaries often play a role in constructing the universe of sites. The number of sites may be as few as three or four to as many as a thousand or more.

When analyzing day trips, one usually defines a cutoff distance for sites to include in the choice set, such as all sites within 200 miles of a person's home (Andrews 1996; Parsons and Massey 2003). In this way, choice sets can vary across respondents. Other times, the choice set is simply all sites in a given county or state and every respondent has every site in his or her choice set (Lew et al. 2008). Occasionally there will be only a handful of sites and the sample will be selected such that the sites are in everyone's choice set (Moeltner and Shonkwiler 2005).

Finally, there has been some concern about using researcher-defined versus individual-defined choice sets. The definitions described above are all researcher-defined. Peters et al. (1995), Hicks and Strand (2000), and von Haefen (2008) have argued that people do not consider this many sites in making a choice and suggest an approach using narrower choice sets determined by people in the survey. In their approach, individuals identify sites they consider in site choice, and this makes up the choice set for analysis. Parsons et al. (1999a, b) argue that considered sites simply have higher utility than sites not considered, and belong in the choice set. Researcher-defined choice sets that ignore consideration sets still dominate the literature.

6.3.4 Step 4. Identify Population of Users and Develop a Sampling Strategy

Next, a sampling strategy is developed. Off-site samples are the most popular for RUM models. This involves sampling from the general population or perhaps from a population of pre-identified users (e.g., people holding a fishing license). Sometimes sampling runs into the issue of low participation rates from the general population—the fraction of people participating is low enough that random samples from the general population yield a small sample of actual users. The conventional way of dealing with this problem is to stratify the sample—sampling counties or communities nearer the sites more heavily because they are likely to have higher participation rates per capita. Stratification not only increases the number of participants per capita but also allows for a more even geographic distribution across the sample of users, which can sharpen parameter estimates by ensuring variation of trip cost across the set of sites. Parsons et al. (2009), for example, used a stratified sample in Texas and purposely under sampled Houston to avoid a population dominated by urban residents and oversampled coastal counties to increase the number of users. Stratification need not affect estimation, but it does affect the calculation of mean welfare estimates across the sample (see Deaton, 1997, pp. 67-73, for more on weighting in estimation). After the values are computed, they must be reweighted to reflect the population in the computation of aggregated welfare impacts. In the example above, values from Houston residents needed to be weighted "up," and values from coastal residents needed to be weighted "down."

There is also the issue of defining the extent of the market. What are the boundaries for the sample? In principle, the relevant population includes all users and potential users of the affected sites—persons who use the sites without the changes and who might use the sites with the changes.

One way to capture the market in a day trip analysis is to define the population as all individuals residing within a day's drive of the affected sites. For a popular ocean beach, this might make sense. For a small fishing pond, it will extend too far. Sometimes a preliminary phone or mail survey is used to gauge the relevant market area.

Market definitions for day trips vary. Lew et al. (2008) define the market for day-trip beach use in San Diego County as all residents (presumably adults) residing in the county. Adamowicz et al. (1997) define the market for hunting in Alberta as Alberta residents holding provincial moose licenses. Parsons et al. (2009) define the population of users for a day trip model as anyone residing within 200 miles of a Texas beach. Hindsley et al. (2011) also use 200 miles as a cut off for day trips.

As noted above, analysts will sometimes use on-site samples. Sampling on-site can be challenging. Think about randomly selecting a user on a beach. When and where do you sample? A strategy must be devised. For example, randomly drawing several weekdays and weekends during the season for interviewing and then, once on-site, interviewing every tenth person one encounters is a strategy that attempts to approximate a random sample. Clear entry points, such as gates, help in on-site random sampling.

Consideration must also be given to how one conducts a survey on-site. Interrupting someone as they sleep on the beach or as they put a boat in the water is hardly advisable. Catching respondents at an opportune time with minimum disruption will help response rates while extending common courtesy.

Another consideration is whether or not to sample people as they arrive or depart. The latter has the advantage that respondents know more about the actual recreation experience—catch rate of fish, time spent on the site, activities, and so forth. However, the former may be a more accurate reflection of an individual's perception of the site that drove the actual choice. Finally, which sites are sampled and how frequently they are sampled must be established, documented, and then used in weighting.

6.3.5 Step 5. Variable Selection

Before data collection begins, the analyst selects the variables that will be included in the model. Site utility in Eq. (6.9) includes trip cost (p_n) and a vector of site characteristics (Q_n) for each site. In some cases, it will include interactions with individual characteristics $(\widetilde{Q}_n)$. The vector Q_n includes characteristics that matter to people in making a site choice. This will vary depending on the type of recreation being studied. Fishing might include catch rate of fish, rock climbing might include difficulty of the climb, boating might include presence of boat ramps, and so forth. It is difficult to make a generic list, but the following are some common attributes that occur in appear RUM models:

- Size of site: length of beach, acres of forest, number of trails.
- Accessibility of site: boat ramp, access by 4-wheel drive only, remoteness.
- Environmental quality: water quality measures, cleanliness.
- Park: federal, state, or local park.

- On-site activities: camping, activity center, surfing conditions.
- Wildlife abundance: catch rate of fish, presence of special species.
- Character of nearby area: shopping nearby, industry nearby.
- Special features: casino, historic significance, amusement park.
- Regulations: bag rates, consumption advisories.

The no-trip utility in Eq. (6.17) or the participation function in the linked model in Eq. (6.19) usually includes a vector of individual characteristics that govern how often and whether a person makes a trip. Some common user characteristics include:

- Family size/number of children.
- Age.
- Gender.
- Urban/rural residence.
- Occupation.
- Level of education.

Family size or composition, for example, may matter for beach recreation where families with young children may be more likely to participate. Hunting is more popular among men than women; bird-watching is more popular among older people; and fishing is more popular among rural residents.

It is common to see attitudinal measures and other participant-specific covariates believed to pick up "intensity" of use in the participation stage of a RUM model. Some caution is warranted here. These variables are likely to have some degree of endogeneity; unobserved characteristics influencing number of trips taken may also influence attitudes and intensity of interest. Attitudinal variables might include answers to response data based on a question like "Would you consider yourself an advocate for the environment?" Club membership is a popular variable also (e.g., belonging to a fishing or hunting club or an environmental group). Measures of experience are also used (e.g., number of years a person has been rock climbing or a self-evaluation of level of experience/expertise based on the response to a question like, "How would you rate your level of rock climbing skill? Novice, intermediate, or expert?"). Again, the problem with all of these variables is that they are likely to be endogenous, and whether the analyst is measuring magnitude of causality in the TCM is questionable. Furthermore, if these variables are correlated with exogenous variables in the model, the parameter estimates on the exogenous variables will be inconsistent (Cameron and Trivedi 2005, p. 92).

In some cases, individual characteristics enter the site choice model as interaction terms to capture preference heterogeneity ($\widetilde{Z}_n$ in Eq. (6.8)).[9] As explained earlier, this happens when one believes that a characteristic affects people differently. For example, the presence of a boat ramp may matter to people who trailer a

---

[9]Another way of accounting for preference heterogeneity is to estimate a Latent Class model, wherein people are sorted into a finite set of TCMs, each with its own set of parameters usually sorted by individual characteristics (Boxall and Adamowicz 2002).

boat but may not matter to people who keep a boat at a marina. Vehicle access on a beach may matter to people who surf fish but not people who are sunbathing.

If one is interested in site closure only, gathering detailed data on all attributes is unnecessary. A set of alternative specific constants for each site is sufficient. In this case, each site has its own constant term (one must be dropped for identification). This will, in effect, perfectly differentiate sites by popularity and gives accurate site-closure estimates. Murdock (2006) argues to include alternative specific constants for all sites, along with characteristic data in any analysis to avoid biases due to unobserved site characteristics.

For good examples of differences in variable selection by recreation types, see Siderelis et al. (1995) for boating, Shaw and Jakus (1996) for rock climbing, Kinnell et al. (2006) for urban parks, Hunt et al. (2007) for fishing, Morey (1981) for skiing, Hynes et al. (2007) for whitewater kayaking, Parsons and Massey (2003) for beach use, and Adamowicz et al. (1997) for hunting.

6.3.6 Step 6. Gather Site Characteristic Data

The next step is to gather site characteristic data. The typical sources are the state and federal agencies responsible for managing the resource. These agencies may have data on indicators of environmental quality and physical measures, such as acres, and sometimes, percent land cover in different uses.

Other sources of data are tourist bureaus, recreation clubs and associations (e.g., hunting or fishing), university researchers, government scientists, and newspapers and magazines. Some cases need fieldwork where the analyst constructs the primary data independently or interviews knowledgeable people (see Leggett et al. (2015) for a good of example of fieldwork used to measure marine debris). Even in cases where variables are not constructed by field observation, such visits can confirm or amend existing site and variable definitions. Online maps are also useful for site definition and for identifying characteristics such as size, presence of water bodies, parking, access points, etc.

In some circumstances, the site characteristic data are gathered in the survey. For example, individuals may be asked to rate the quality of hunting at each site visited in the current season. Using these responses, the analyst constructs a hunting quality index for each site. One could do the same with catch rate of fish, view amenities, and so on.

There are several problems with this approach. First, the data for each site are usually confined to those who have visited the site. This is likely to bias quality ratings upward. People who visit a site are those who find the site desirable. Second, popular sites have more data than less popular sites. Many sites may have a single visitor or no visitors, which causes an asymmetry in the quality of variable measurement across sites. Third, perceived variables are more difficult to map into actual policy changes.

Another method for working with survey data is to perform an auxiliary regression using the reported measure for the characteristic (such as number of fish caught) as a dependent variable and observable site characteristics (such as lake size, depth, elevation, and presence of regulations) as explanatory variables. The fitted regression is then used to predict characteristic values at each site.[10] This allows for prediction at sites where the reported measure is unavailable and for use of "perceived" measures of quality if the researcher prefers these.

6.3.7 Step 7. Design and Implement the Survey

Next is the design and implementation of the survey (see Chap. 3 for more detail). Begin with an existing set of recreation surveys to get a sense of layout and question format. The survey is usually composed of four key sets of questions to measure the variables appearing in the model:

1. Trip count and location questions.
2. Last trip questions.
3. Stated-preference questions if applicable.
4. Respondent and household characteristic questions related to the activity being modeled.

The trip count questions ask the respondent to report the number of trips taken over a designated time period and the sites visited. These questions may be divided by recreation type (fishing, boating, etc.), by day and overnight trips, and/or by multiple- and single-purpose trips.

There are essentially three approaches for counting trips to the sites.

1. The researcher provides the respondent with a preestablished list of sites. The respondent reviews the list and records the trips over the relevant time period. This approach is the easiest for data handling; the sites are predefined and each respondent's count fits neatly into the definition. This approach also helps recall. The site names serve as a reminder to people and recalling exact site names is not necessary.
2. The second approach is open-ended. People are asked to list all the sites they have visited over the relevant time period, along with a count of trips to each. Once gathered, the data must be reworked such that site definitions are consistent across respondents. This can be time-consuming. Sites often have more

[10]For an example applied to fish catch, see McConnell et al. (1995). For a debate on the validity of this strategy, see Morey and Waldman (1998, 2000), and Train et al. (2000).

than one name, people may use nearby town names, and sites with similar names can lead to confusion. A useful innovation to use in conjunction with the open-ended approach is a map. Respondents are asked to identify each site visited on a map. This approach eases the final assembly of the data and avoids any confusion about the name and location of the site. It also has the advantage of aiding in finer site definitions depending on individuals' use of sites. For example, if there is heavy visitation to a lake that seems divided by trips to the northern and southern portions of the site, one might opt for dividing the lake into two sites.

3. A third approach is to work with last trip data alone. In this case, the researcher gathers information only on the last site visited—which site, how many trips were taken to that site over the season, and how many were taken to all sites in total? A basic RUM model can be estimated with these data. The survey is less complex, and recall is a smaller issue. This comes at the expense of fewer trips and hence, there is less data for the final analysis. In principle, the analysis could be done with any single site visited during the season. For example, one could ask for the site visited closest to a specific date where the date is varied to assure that trips are spread across the season.

Regardless of the approach chosen above, detailed data is often gathered for only the last trip taken—time spent on-site, number of people sharing travel expenses, other expenses incurred, information on the trip experience such as number of fish caught, etc. These data—used to construct explanatory variables in the demand model—are usually gathered for the last trip only because gathering them for each trip over the season can lengthen a survey considerably, and the information is difficult for respondents to recall for every trip. Sometimes analysts will gather data on the last two or three trips to expand the data set. Another alternative is to use a "typical trip." In this case, individuals are asked to report on a self-identified typical trip. Finally, if the number is not too large and the recall period is short, detailed information might be gathered on each trip.

If one is combining the revealed preference data on trips with stated-preference data, questions that ask how site choices and number of trips change under different conditions of quality are included. This effectively renders two sets of trip data: one for the revealed preference or actual trips and another for the stated-preference trips (Jakus et al. 2011). There are two approaches for expressing stated-preference responses: retrospective or prospective. In the retrospective case, the analyst asks the respondent how he or she would have done things differently if conditions had been worse or better. In the prospective case, the analysts first asks how many trips the respondent plans to take next year and then how that number would differ if conditions were worse or better. The advantage of the former is that it places an individual in the actual choice situation and so may have more realism and grounding with actual behavior (Parsons et al. 2013). The latter has the advantage of giving the analyst stated-preference data for current conditions in the form of future trips, which can be used (in principle) to correct for stated-preference bias that might be present in the stated-preference data associated with the change in trips due to a change in conditions (Whitehead et al. 2008).

Usually, in the final section of the survey, the researcher asks respondent and household characteristic questions for the elements in Z_n. These include age, family composition, etc., and income and location of the respondent's hometown, which are required to estimate trip cost. In some cases, analysts will add attitudinal questions as well.

While there is no need or room to repeat the principles of good survey research here, there are several issues of special concern in travel cost surveys that deserve some attention: trip recall and trip categorization.

When people are asked to report their number of trips, the analyst assumes they will remember how many were taken. Because the model calls for a count of trips over a season, respondents are often asked to recall the number of trips taken over many months or even a year. But how accurate is an individual's recall of past trips? And will approximations be valid?

The evidence on how large the error may be is slim, but it is hard to deny the potential. One approach to improve recall is to survey at several intervals over the season, asking respondents to report trips taken only over the preceding month or so. Alternatively, one might ask respondents to keep a trip log during a season. While sensible, this approach can raise survey costs considerably and lower response rates through respondent attrition over the season. Fisher et al. (1991) provides the most systematic study of recall bias in recreation use surveys. For hunting and fishing, they find that the number of reported trips increases with the length of the recall period. They considered periods as short as one month and as long as a year in the context of the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. Their results certainly point to shortening the recall period to improve accuracy. Even going from three months to one month made a difference in recall.

Off-site surveys are usually conducted immediately following a season. On-site surveys are done within a season and are truncated because the respondent can report only the number of trips taken to date; the balance of the season is unknown. Two approaches are used to fill in the data for the end of the season: (1) ask respondents to estimate the number of trips they plan to take or (2) predict the number of trips based on reported trip behavior (Egan and Herriges 2006; Parsons et al. 2013). On-site samples where individuals are recruited for an end-of-the-season mail or Internet survey, will, of course, provide a complete trip count because the survey itself is completed at the end of the season.

As mentioned in Step 2, there is often more than one type of recreation at the site. If so, the survey may be designed for many recreation types. The most common survey strategy is to proceed use by use in the questioning. For example, one might begin by asking, "How many day trips did you take to the site for the primary purpose of fishing?" and follow with similar questions for swimming, boating, etc.

There are shortcuts that reduce the length of the survey. One might simply ask people to report, "How many trips did you take to the site for purposes of recreation?" And then, "What is your primary recreation use of the lake?" This avoids questioning about each recreation use. People are then classified by their primary recreation activity. Different models of lake visitation, for example, are estimated for people who primarily fish, who primarily go boating, and so on. There is some mixing of trip activity types within models in this case; however, if people have a single, dominant use for a site, this strategy is effective.

It is also useful to identify side trips and trips originating from somewhere other than the resident's hometown. For example, if a person is on a business trip and takes a whale-watching tour while on the trip, the whale-watching tour is a side trip. The trip cost from the individual's home to the site would overstate the real marginal cost of the trip. It is easiest to identify and delete side trips from the analysis by clearly defining a day trip as being to and from one's primary residence when the trip count is being made.

A similar issue arises when an individual owns a cabin or cottage near the recreation site of interest. Suppose the person spends most of the summer at the cottage and makes frequent day trips to the site from that secondary residence. How should these trips be handled in the analysis—as one three-month-long overnight trip from their permanent residence or as many day trips to the site from their cottage? While the former choice should, in principle, embody the individual's entire consumer surplus, it is not amenable to the travel cost model. One might consider two strategies: either drop the observation from the sample and lose a heavy recreation user or include the observation and count the trips as day trips from the cottage. To identify these people, it is necessary to include a question that asks whether or not the individual owns (or rents) property near the recreation site, uses it for extended stays, and makes trips to sites from there.

6.3.8 Step 8. Measure Trip Cost

Once the data are assembled and organized, the trip costs to all sites for each respondent are computed. Trip cost is the sum of the expenses required to make a trip. Typical costs for a day trip include:

- Travel cost.
- Access fees.
- Equipment cost.
- Time cost.[11]

A basic premise underlying the TCM is that these costs, which are the price of a trip, are taken as given to individuals. In fact, all are to some extent "chosen." People choose their vehicle type, which affects travel cost. People choose the type of equipment to use for fishing, which affects trip cost. People choose where to live

[11]Expenses like food and souvenirs are typically excluded because they are not necessary to make the recreation trip possible.

(and may have done so based on proximity to recreation sites), and this affects travel and time cost. And finally, people certainly choose the amount of time they spend on-site for any given trip. The TCM implicitly assumes that the price variation generated by these “chosen” aspects to trip cost can safely be ignored.[12]

The first element in the list, travel cost, is typically computed using a per-mile operating cost times round-trip mileage to the site(s). For example, the American Automobile Association’s 2015 average cost per mile for operating a vehicle (average of all sedans) is about 15 cents ($0.15), which includes the cost of gas, maintenance, and tires. This is close to the incremental cost of a trip. (See Hang et al. (2016) for an argument for excluding derpreciation cost). If the type of vehicle is known, the analyst can differentiate cost by type (e.g., SUV versus compact car). The round-trip distance to the sites is usually calculated using a software package such PC*Miler. Tolls, if any, are added to operating costs.

Because travel costs may be shared by several people, efforts are sometimes made to apportion the costs. For example, the researcher might ask respondents to report the number of people sharing travel costs on their last trip and divide the cost equally. Or, the analyst might ask directly for an individual’s share of the cost.

If sites have an access fee, that fee is included in the trip cost. Sometimes sites will offer annual or weekly passes, senior discounts, or free admission for children. Making adjustments for seniors and children is easy, but accounting for discounts is more difficult and usually ignored. Typically, the daily fee is used. Hunting and fishing licenses can be expensive and are necessary to make a trip possible, but they are a fixed, not marginal, cost per trip. A researcher might consider using the cost of a license only in the participation stage of the single-site model (Eq. 6.6a) or seasonal portion of a RUM model. This would capture its influence on participation and treat it as a fixed cost.

Equipment costs vary by type of recreation. For fishing, one needs bait, tackle, a rod, and sometimes the use of a boat. For beach use, there may be chairs, umbrellas, surfboards, and so on. For bird-watching, there are binoculars and film. For an item like bait, the cost is simply the market price of bait for a day of fishing. For durable goods, an imputed rent is needed. If one rents or charters a boat for fishing, the cost is simply the fee; if one owns a boat, the rent or cost of its use needs to be imputed. One approach for imputing such costs is to use the rental fee for comparable services. This is, no doubt, an overstatement of the cost, but it is easier to obtain. However, equipment costs are often excluded from the trip cost estimate because they are difficult to estimate and are generally negligible portions of trip cost when viewed in terms of imputed rent. Also, if equipment costs are constant across the sample of users (or nearly so), they can be safely ignored without biasing the trip cost coefficient.

---

[12]Several studies have considered endogenous trip costs. Parsons (1991) analyzes endogenous residence choice and Baerenklau (2010) has a nice follow-up with some contrasting results. Bell and Strand (2003) let choice of route be endogenous. McConnell (1992), Berman and Kim (1999), Offenback and Goodwin (1994) analyze endogenous on-site time.

An alternative for estimating travel expense, access fees, and equipment cost is to ask individuals to report these expenses on their last trip to the site. The advantages of this approach are that it uses perceived cost information, and the researcher need not construct the cost estimates. Because individuals base trip decisions on perceptions of cost, which may diverge from actual costs, the respondent-reported estimate is somewhat compelling. Objective estimates based on researcher computation are still most common.

The most difficult issue in computing trip cost, and certainly the one that has received the most attention, is estimating the time cost of the trip. Time spent traveling to and from the site and time spent on the site constitutes time that could have been devoted to other endeavors. The value of those lost opportunities is the time cost of the trip. Time cost often accounts for a sizable portion of the total trip cost.

In most applications, the estimate of time cost is related to a person's wage. This relationship has a theoretical basis as long as the individual has a flexible working arrangement and can substitute work time for leisure time at the margin. Under such conditions, in theory, an individual increases the number of hours worked until the wage at the margin is equal to the value of an hour of leisure. Multiplying the hourly wage by travel and on-site time results is an estimate of time cost. Unfortunately, this simple model breaks down for many individuals (Bockstael et al. 1988). The simple leisure/work trade-off does not apply to individuals working a fixed 40-h-per-week job for a salary. These people do not have the flexibility to shift time in and out of work in exchange for leisure. The trade-off is also implausible for retired persons, homemakers, students, and unemployed persons.

Despite the theoretical difficulty of extrapolating the flexible leisure/work model to many persons, the most commonly used approach to value time is still wage-based. For people with fixed work schedules, most studies impute an hourly wage using annual income. Reported annual income is divided by the number of hours worked in a year—usually a number in the range 2,000 to 2,080—giving an imputed wage. For people with an hourly wage, if it is reported, that wage is used. Then, it is typical to use one-third of the wage as the opportunity cost of leisure time, though occasionally 25% or even the full wage is used. In a quick review of the literature since 2000, about half of the studies used one-third of the hourly wage. The other half was about evenly divided above and below 33%. The practice of using a fraction of the wage stems from the early transportation literature wherein analysts had imputed the time cost in empirical travel studies at less than traveler's full wage. Some analysts also argue that the trip itself may impart some enjoyment to individuals, and one way of accounting for this is by discounting the value of time—the loss is diminished somewhat by the trip itself being pleasurable.

Another approach for handling time valuation for those with a fixed salary is to impute a wage using a simple hedonic regression over the subset of individuals in the sample earning an hourly wage (Smith et al. 1983). Wage is regressed on income and a vector of individual characteristics such as age, gender, and education. The fitted regression is then used to impute a wage for people who do not earn wages. There are also approaches for inferring values of time internally within the TCM. For example, if one enters separate terms for travel time and out-of-pocket costs in the TCM, the ratio of the coefficients is a measure of the value of time (McConnell and Strand 1981), and one can use the out-of-pocket cost coefficient directly in valuation having controlled for time. The difficulty with this approach is the high colinearity between time and out-of-pocket costs.

It may also be possible to divide the sample by those who are able to adjust work hours and those who cannot (McConnell and Strand 1994). In this case, the analyst uses wage to value time and adds it to out-of-pocket costs to arrive at trip costs for the first group, while time and out-of-pocket costs are entered separately for the second group on the grounds that time costs are unknown. Similarly, there are approaches where one identifies whether a person is over-, under- or fully employed using survey data and then estimates a shadow wage equation from which values are inferred (Feather and Shaw 1999).

On-site time is usually excluded from the trip cost computation. McConnell (1992) presents a theoretical argument justifying its exclusion. He treats time on-site and the trip as separate choices, each with its own "price." In this way, the quantity of time on-site is removed from the trip cost calculation because it is chosen in the model. The TCM then requires a price for on-site time, but this can be ignored if it is uncorrelated with the other variables in the model. (Also, see the studies mentioned in Footnote 12 about handling on-site time.)

In summary, the most common form for "price" used in travel cost studies is

$$p_n = \left\{ \left(0.33 \cdot \left(\frac{y_n}{2{,}040} \right) \cdot t_n \right) + (c_n \cdot d_n) \right\} + \text{fee}_n + \text{tolls}_n + \text{other}_n, \tag{6.20}$$

where y_n is individual annual income, t_n is the round-trip travel time to the site, c_n is the cost per unit mile, d_n is the round-trip distance to the site in miles, fee_n is a site access fee, tolls_n is the cost of tolls, and other_n is other expenses required to make the recreation experience possible (often set to zero). If a person is a wage earner, that wage is used instead of $\frac{y_n}{2{,}040}$ in the above equation. Again, the calculation is made for every site for each respondent.

6.3.9 Step 9. Estimate Model

The next step is to estimate the model, which is covered in Sect. 6.3. Many software packages exist for limited dependent variable and discrete choice models. LIMDEP, STATA, MATLAB, R, and GAUSS are popular choices.

6.3.10 Step 10. Report Study Results

The final step is reporting the study results. This should include a discussion of the method, summary of the data, model, estimation results, and values. One should look at published articles for examples. It is common to report values in per-trip, per-choice occasion, and seasonal terms. The more general the presentation of the results, the more likely they will be useful for benefits transfer (see Chap. 11). One should also think about users and policy applications when reporting results. Summary statistics for data are often useful to put results in context. Ranges of estimates to capture uncertain and different scenarios can be useful. Comparisons

Table 6.2 Steps to estimate beach use RUM model

Step 1	Application covers closure of 10 southernmost beaches in New Jersey (access value) and the gain from widening the 10 southernmost beaches in New Jersey (quality change)
Step 2	All beach recreation is aggregated into a single use. The analysis was for one year, 2005
Step 3	The choice set includes 66 ocean beaches in the states of New Jersey, Delaware, and Maryland. The beaches are defined using the political boundaries of beach communities and parks
Step 4	The analysis considers day trips for anyone 18 years or older living within a four-hour drive of the coast. An internet sample is used. Coastal states are sampled more heavily than noncoastal. Because the analysis covered day trips and people were asked to report primary purpose trips, all trips are assumed to be single-purpose (see Sect. 6.5.1 for a discussion of single-versus multiple-purpose trips)
Step 5	The RUM model is specified with the following beach characteristics: trip cost, degree of development, park, private, boardwalk, beach width, facilities, plus more. The individual characteristic data included occupation, education, family composition, ownership of surf fishing equipment, and other household characteristics. See Table 6.3 for a complete list of variables
Step 6	The site characteristic data are gathered from various sources: state departments of natural resources (including interviews with experts in those agencies), field trips to sites, interviews with a scientist familiar with the physical characteristics of the New Jersey beaches, tourist guides, maps, newspapers, and websites
Step 7	Nearly 2,000 people completed the survey. About 28% had taken at least one trip to an ocean beach in 2005. Individuals were asked to report day, short overnight, long overnight, extended stay, and side trips separately. The survey first asked people to report the states in which they had visited an ocean beach in 2005 and then showed them a list of beaches in the relevant states. Maps of the shoreline were shown for each state
Step 8	Trip cost was measured as the sum of travel expense, time expense, and beach fees. Many trips to New Jersey beaches are via toll roads, and on some routes a ferry is used to cross the mouth of the Delaware Bay. After the shortest route was determined, toll roads were identified and their cost computed. Many New Jersey beaches have fees. A per-day fee that was published in a local newspaper for each beach was used. Time costs were estimated as a wage proxy times

(continued)

Table 6.2 (continued)

	round-trip travel time. The wage proxy was annual household income divided by 2,080
Step 9	Table 6.3 shows the results for a mixed logit RUM model and linked negative binomial model. The site characteristics are variables appearing in the site utility (see Eq. (6.9)). Stefanova (2009) also included a set of regional constants to pick up effects shared by beaches within these regions. These are the variables in the table beginning with Barnegat and ending with Maryland. Because these are mixed logit results, the estimates include means and standard deviations for each variable (see Eq. (6.16)). The coefficients on the individual-specific characteristics are for the trip frequency model As shown, the site characteristics that significantly increase a site's day-trip utility are boardwalk, amusements, park, beach width, and presence of high-rise. The characteristics that decrease a site's day-trip utility with statistical significance are private and vehicle access. Several of the variables have large standard deviations suggesting substitution across sites along the lines of these attributes is higher. Amusements, parks, and Atlantic City, for example, show wide variation relative to their mean values. The individual characteristics that increase a person's probability of taking a trip and the number of trips taken in the trip frequency stage are age, working full-time, having flexible work hours, and owning vacation property. The coefficient on the log-sum term is positive as expected, meaning trips increase as the expected value of a trip increases
Step 10	The site access values for the closure of the 10 southernmost New Jersey beaches are shown in Table 6.4 for three models: multinomial logit, mixed logit, and multinomial with alternative specific constants only. All trip models use the linked model for trip frequency. Per-trip loss, per-season loss, loss-to-trips ratio, and aggregate seasonal loss are shown[a]. The estimates are stable across the models. The simple multinomial logit with alternative specific constants predicts about the same access values as the logit models with detailed site characteristics. Aggregate annual losses for day-trip beach use to beaches in the Cape May area is about \$435 million in 2005 dollars The results for the gain in beach width are also shown in Table 6.4, in this case for the mixed logit only and for three scenarios: widening all beaches to 100, 150, and 175 feet. Because some beaches are already at or wider than these widths, only some beaches are affected by each policy increment. One beach is affected at 100 feet, three at 150 feet, and four at 175 feet. As expected, the gains are much lower than the access value losses, and they roughly double as one moves from one widening scenario to the next. The aggregate annual value for the biggest improvement is \$11.4 million

[a]The loss-to-trips ratio is total access value for the sample divided by the number of trips taken to the site or sites. It is commonly used in damage assessment where one knows the total number of displaced trips from a site or sites and needs an estimate of loss per trip. It is different from a conventional average per-trip value in a RUM model (Eq. (6.13)). That value gives the *average of the change in the expected value of a recreation trip to everyone in the sample* for a change in access or quality to some set of sites affected by a policy. Everyone in the sample is included, whether or not a person actually visited the affected sites. In a RUM model, everyone will realize some effect from a closure or quality change at every site because everyone has some nonzero probability of visiting every site. The loss-to-trips ratio is instead a unit value for each of the trips taken specifically to the site where the change in access or quality is occurring. With the per-trip value, one multiplies by total trips to all sites. With loss-to-trips, one multiplies only by trips taken to the affected sites

Table 6.3 Mid-Atlantic ocean beach RUM model

Mixed logit site choice			
Variable	**Variable definition**	**Mean parameter estimate (*t*-stat)**	**Std. deviation parameter estimate (*t*-stat)**
Travel cost	Travel plus time cost	−0.046 (−40.02)	–
Boardwalk	Boardwalk present = 1	0.199 (3.72)	0.088 (0.28)
Amusements	Amusements nearby = 1	0.774 (10.52)	1.357 (4.27)
Private	Private or limited beach access = 1	−0.520 (−9.26)	0.025 (0.08)
Park	State or federal park = 1	0.585 (2.60)	1.400 (5.20)
Log (width)	Log of width of beach in miles	0.191 (6.17)	0.002 (0.03)
Atlantic city	Atlantic city = 1	−0.183 (−0.67)	2.464 (8.77)
Surfing	Good surfing = 1	−0.121 (−1.11)	1.153 (2.67)
High-rise	High-rises present on beach = 1	0.480 (12.66)	0.014 (0.14)
Park within	Park located within beach = 1	0.137 (1.40)	0.555 (1.77)
Facilities	Bathhouse, restrooms present = 1	−0.066 (−1.47)	0.044 (0.23)
Vehicle access	Vehicle access on beach = 1	−0.989 (−4.45)	0.902 (2.75)
Barnegat	Barnegat = 1	−0.138 (−2.32)	0.010 (0.08)
Long Beach Island	Long Beach Island = 1	−0.156 (−1.65)	0.058 (0.30)
Atlantic	Atlantic = 1	0.527 (5.23)	0.077 (0.36)
Cape May	Cape May = 1	1.306 (13.07)	0.135 (0.53)
Delaware	Delaware = 1	2.430 (17.47)	1.065 (4.49)
Maryland	Maryland = 1	2.771 (13.78)	1.326 (4.56)
Log length	Log of length of beach in miles	0.150 (6.13)	–
Day trips		7,791	
Participants		560	
Log likelihood		−23,264.04	

Negative binomial trip frequency	
Variable	**Coefficients**
Intercept	−2.03
Log sum/β_y	0.02*
Log age	0.52*
Work full-time	0.90*
Own vacation home	1.55*
College graduate	−0.14

(continued)

Table 6.3 (continued)

Negative binomial trip frequency	
Variable	**Coefficients**
Work flexible hours	0.40*
Have children 0 to 12 years old	0.12
Dispersion parameter	6.68*
N	**1,966**
Log likelihood	**17,793.4**

*Indicates statistical significance with 95% confidence

Table 6.4 Mid-Atlantic ocean beach welfare estimates

Access value for loss of all 10 Cape May beaches				
Model	**Per-trip loss, per person ($)**	**Per-season loss, per person ($)**	**Loss-to-trips ratio* ($)**	**Aggregate seasonal loss ($ millions)**
Multinomial logit (ML)**	5.45	13.68	27.98	440
ML-ASC only**	5.38	13.48	27.15	434
Mixed logit	5.48	13.46	25.81	434
Beach width value for three widening scenarios for all 10 Cape May beaches				
Widen all beaches to: (feet)	**Per-trip gain, per person ($)**	**Per-season gain, per person ($)**	**Gain-to-trips ratio* ($)**	**Aggregate seasonal loss ($ million)**
100	0.03	0.10	0.18	3.10
150	0.06	0.17	0.33	5.50
175	0.13	0.35	0.68	11.40

*See footnote a in Table 6.3

**ML and ML-ASC parameter estimates are not reported here

with similar applications can also put the numbers in context and provide some validity. In some cases, the results may be reported for specific policy purposes.

6.4 An Application to Beach Use Using the RUM Model

This section presents an application to beach use following the steps in Table 6.1. It is based on work by Stefanova (2009) and is shown stepwise in Table 6.2. The beaches are in the Mid-Atlantic region of the United States.

6.5 Other Considerations

This section addresses some other consideration in travel cost models: multiple-purpose and overnight trips, intertemporal substitution, and congestion. These apply to all TCM types.

6.5.1 *Multiple-Purpose and Overnight Trips*

Sometimes the purposes of a trip will extend beyond recreation at the site. For example, a person may visit family and friends, go shopping, or even visit more than one site on a recreation outing. In these instances, the trip is producing (i.e., the trip cost is buying) more than single-purpose recreation, and it is no longer clear that the simple travel cost paradigm applies. For this reason, researchers often confine their analysis to day trips where multiple purposes are less likely to occur. This is done by either (i) identifying day trips in a survey question or (ii) focusing on trips made within a day's drive from a person's home and assuming that these will largely be day trips. Another strategy is to ask people to report only trips made primarily for the purpose of recreation and treat all of these trips as single purpose.[13]

The analysis gets more complicated if overnight trips are included. First, there is problem of multiple purposes. But, even if the researcher can identify single-purpose overnights trips, they are problematic. There are more costs to estimate (e.g., lodging at all sites). Length of stay can vary significantly over the sample (e.g., some people stay one night, others for two weeks). The relevant choice set is likely to be considerably larger. For example, for a household in the United States, the set of substitutes for a weeklong beach vacation may include all beaches in the United States and beyond. Also, if people use long trips as getaways, nearby sites with a low trip cost may be undesirable. Greater trip cost then, at least over some range, would be viewed as a positive attribute, complicating "price" in the simple TCM. Finally, many overnight trips will be multiple-purpose/multiple-site excursions, wherein the individual transits from one site to the next, which obviously strains the TCM paradigm.

There are many applications in the literature where overnight trips are included (Bin et al. 2005; Parsons et al. 2013; Grijalva et al. 2002). In most cases, it is common to assume that all the variation in trip costs is due to differences in trip distances across the sample. If so, lodging cost can be ignored. It is subsumed into

[13]Even if some trips are multiple-purpose, Parsons and Wilson (1997) show that multiple-purpose trips where all other purposes are incidental, can be treated as single-purpose trips. If people say trips are primarily for the purpose of recreation, one may be able to safely assume all other purposes are incidental.

the constant term of the trip frequency portion of the model. There are a few studies that attempt to model trip length (Shaw and Ozog 1999; Yeh et al. 2006; Karou 1995; Hoehn et al. 1996). One difficulty with this approach is configuring the choice occasion such that a person may choose to take an overnight trip, a day trip, or several day trips on a single occasion.

In cases where individuals visit multiple sites on a single trip, one promising approach is to redefine a site such that there are "single-site" and "multiple-site" alternatives, and then proceed with the logic of the TCM (Mendelsohn et al. 1992). Trip costs are recalculated for a "multiple-site" alternative by accounting for the costs of reaching all the sites on one trip. In a similar vein, one might cast the RUM model as a portfolio choice problem (Tay et al. 1996). Instead of individual sites, alternatives are defined as a portfolio of sites. Each site is represented by a dummy variable in a portfolio utility, and trip cost is the cost of visiting the entire portfolio. People then choose portfolios. Welfare changes would be computed for losses of one or more sites from the portfolios. Because the number of possible portfolios is likely to be large, estimation by randomly drawing portfolios probably may make sense.

Another approach in the RUM context is to handle "other purposes" as an attribute of a site. For example, nearby shopping may be a variable in a RUM model of beach use. This, in effect, expands the nature of the recreation experience.

6.5.2 Intertemporal Substitution

The bulk of the literature and most of the active research on TCMs ignores the dynamic aspect of decision-making, but it is hard to deny its importance. Dynamics allow people to substitute sites over time and allow experiences early in the season (such as a good catch rate of fish) to affect choice of site and/or number of trips taken later in the season. Individuals may even base current site choices on expectations about future trips.

In principle, the Repeated RUM model is set up for just such an analysis because it considers an individual's trip choice day by day over a season. Nevertheless, few applications consider interdependence in this framework. The typical analysis treats each trip choice as independent of the previous and upcoming choices and takes no account of temporal characteristics (such as weather, day of week, and so forth). There are two good reasons for this. First, the data are more difficult to collect. To gather trip data by date of trip usually requires a diary by respondents for recall of dates to be accurate. This means repeating survey administration (perhaps monthly throughout the season) or continual reminders to complete a diary sent early in the season. This increases the cost of the survey and leads to sample attrition. Second, there is an inherent endogeneity in trip choice over time. Unobserved factors affecting trips in period are no doubt present in periods $t-1$ and $t+1$. If so, this feedback needs to be dealt with by purging the endogeneity present in explanatory variables with historical or future content (most notably the lag of past trips to the site used as explanatory variables).

There have been several efforts to build time interdependence into TCMs. As just noted, one way is to use a measure of past trips to a site as an explanatory variable in a travel cost model (McConnell et al. 1990; Adamowicz 1994; Moeltner and Englin 2004). A positive coefficient here implies "habit formation" and a negative coefficient implies "variety seeking." McConnell et al.'s (1990) analysis is in a single-site setting and is notable for attention to the need of instrumental variables to identify the effect of past trips on current behavior.

Adamowicz's (1994) and Moeltner and Englin's (2004) analyses are in a RUM setting. Both use the number of past trips as explanatory variables. Moeltner and Enlgin (2004) also include the consecutive number of times a site was chosen up to the current period—uninterrupted by any visits to other sites—as a measure of brand loyalty. Provencher and Bishop (1997) estimate a fully dynamic model where choices over the season are the result of solving a dynamic programing problem—past and future trips are integrated into current trip decisions. Theirs is a single-site discrete choice model (go/don't go each day of the season). Finally, Swait et al. (2004) link past trips to current behavior using a "meta" site utility function that links a season's worth of site utility functions into a single expression and optimization over a season.

There have also been some efforts that combine stated-preference data with trip choice data to infer intertemporal effects. For example, suppose people were asked what they would do if the site they last visited had been closed. If the response format allowed, among other things, for people report taking trips later in the season to make up for the lost trip, there is some empirical data on time interdependence (Parsons and Stefanova 2011). Finally, single-site models estimate diminishing marginal utility of trips to individual sites within a season and, at least in this way, if it exists, implicitly capture the extent of habit-forming versus variety-seeking behavior in the sample.

6.5.3 Congestion

Given the growth in population and income and the decline in transportation cost over time, recreation sites see more use. In some cases, congestion becomes an issue. While the theory of incorporating congestion into TCMs is well understood (Freeman 2003) as is the idea of efficient or optimal congestion at a site (Fisher and Krutilla 1972), it is difficult to incorporate its effect empirically in a TCM.

The difficulty is captured in a famous Yogi Berra quote that a popular restaurant "is so crowded that nobody goes there anymore." Observing many people at a site is evidence of its desirability and, hence, its high probability of visitation. At the same time, it may have gotten so popular that visitation is actually somewhat less than it would have otherwise been but for congestion. How does one tease out the latter effect? An obvious start is to put some measure of congestion on the right-hand side of a single-site or RUM model. But almost any measure one considers is correlated with excluded unobserved variables that influence individual demand—the same factors that influence demand (making a site desirable) influence congestion. There is an inherent endogeneity problem and sorting out the partial effect of congestion, without some clever instruments, is not possible.

There are a few applications in the TCM literature addressing congestion using theories of congestion equilibrium and instrumental variables to identify congestion effects (Timmins and Murdock 2007; Boxall et al. 2001). Timmins and Murdock's (2007) approach is based on an equilibrium model of sorting in a RUM context and is applied to fishing in Wisconsin. Boxall et al.'s (2001) application involves a congestion forecasting model in combination with a RUM model of site choice for wilderness recreation. An alternative strategy is to add a stated-preference question to a survey, wherein one introduces hypothetical levels of congestion (McConnell 1977; Boxall et al. 2003).

6.6 Conclusions

The travel cost model has advanced in ways Harold Hotelling would never have imaged. It has certainly stood the test of time. The demand for its use in legal and policy settings seems to grow steadily, and academics have vigorously explored avenues for improvement. I will not venture a guess as to the direction of future research but will highlight some of the "soft spots" in TCM where improvement is still needed.

1. While there have been numerous efforts at valuing the time in the TCM context, in most applications, simple wage analogies are still used. More realistic measures are needed.
2. There is considerable room for improvement in overnight trip modeling. This would be useful in many settings, including beach use and ecotourism.
3. Models to deal with multiple-purpose trips in a realistic way would allow analysts to consider recreation use in areas where excursions are the norm (e.g., trips to parks out west in the United States).
4. Intuitively, most economists would agree that intertemporal substitution is an important part of recreation behavior, yet all of the modeling efforts along these lines are narrow specialty applications that have seen limited use in the broader, more general literature. Practical, realistic models are needed to deal with this issue.
5. The Kuhn-Tucker model, which integrates site choice and trip frequency using seasonal data, has been on the cusp of taking off for more than a decade but still has not seen widespread use. Lift off in the next decade would be desirable.
6. The expansion of models that include a choice of features, such as lodging, access points, target fish species, time of week, and so on, bring a dimension to

the TCM where individuals are likely to be making important welfare-revealing choices. More work in this area would be beneficial.

7. Given the agreed importance of recall in recreation demand, it is surprising how little we actually know about the error introduced by recall bias. Targeted studies exploring the extent of bias would make a valuable contribution.
8. While much attention is given to the measurement of time in the TCM, how one measures out-of-pocket travel cost (cost per mile traveled) varies dramatically across studies. Some exploration narrowing the values used would help.
9. Finally, integration with stated-preference studies has been expanding and is likely to be an area where large gains can continue to be made.

With work in these and other areas, Hotelling's basic concept of the TCM should continue to flourish.

References

Adamowicz, W. L. (1994). Habit formation and variety seeking in a discrete choice model of recreation demand. Journal of Agricultural and Resource Economics, 19, 19-31.

Adamowicz, W. L., Swait, J., Boxall, P., Louviere, J. & Williams, M. (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Journal of Environmental Economics and Management, 32, 65-84.

Andrews, T. (1996). A discrete choice model of recreational trout angler benefits in Pennsylvania. Unpublished manuscript, Department of Economics, West Chester University, West Chester, PA.

Awondo, S. N., Egan, K. J. & Dwyer, D. F. (2011). Increasing beach recreation benefits by using wetlands to reduce contamination. Marine Resource Economics, 26, 1-15.

Baerenklau, K. A. (2010). A latent class approach to modeling endogenous spatial sorting in zonal recreation demand models. Land Economics, 86, 800-816.

Bell, K. P. & Strand, I. E. (2003). Reconciling models of recreational route and site choices. Land Economics, 79, 440-454.

Ben-Akiva, M. & Lerman, S. R. (1985). Discrete choice analysis: Theory and application to travel demand. In Marvin Manheim (Series Ed.), MIT Press Series: Vol. 9. Transportation studies. Cambridge, MA: MIT Press.

Berman, M. D. & Kim, H. J. (1999). Endogenous on-site time in the recreation demand model. Land Economics, 75, 603-619.

Bin, O., Landry, C. E., Ellis, C. L. & Vogelsong, H. (2005). Some consumer surplus estimates for North Carolina beaches. Marine Resource Economics, 20 (2), 145-161.

Bockstael, N. E., Hanemann, W. M. & Kling, C. L. (1987). Estimating the value of water quality improvements in a recreational demand framework. Water Resources Research, 23, 951-960.

Bockstael, N. E., Hanemann, W. M. & Strand, I. E. (1984). Measuring the benefits of water quality improvements using recreation demand models. Report presented to the U.S. Environmental Protection Agency. College Park: University of Maryland.

Bockstael, N. E., McConnell, K. E. & Strand, I. E. (1988). Benefits from improvements in Chesapeake Bay water quality. Report presented to the U.S. Environmental Protection Agency. College Park: University of Maryland.

Boxall, P. C. & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: A latent class approach. Environmental and Resource Economics, 23, 421-446.

Boxall, P., Hauer, G. & Adamowicz, W. (2001). Modeling congestion as a form of interdependence in random utility models (Staff paper 05-01). Department of Rural Economy, University of Alberta, Edmonton, Alberta, Canada.

Boxall, P., Rollins, K. & Englin, J. (2003). Heterogeneous preferences for congestion during a wilderness experience. Resource and Energy Economics, 25, 177-195.

Breffle, W. S. & Morey, E. R. (2000). Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of Atlantic salmon fishing. Marine Resource Economics, 15, 1-20.

Brown, G. Jr. & Mendelsohn, R. (1984). The hedonic travel cost method. Review of Economics and Statistics, 66, 427-433.

Brown, W. G. & Nawas, F. (1973). Impact of aggregation on the estimation of outdoor recreation demand functions. American Journal of Agricultural Economics, 55, 246-249.

Burt, O. R. & Brewer, D. (1971). Estimation of net social benefits from outdoor recreation. Econometrica, 39, 813-827.

Cameron, A. C. & Trivedi, P. K. (2005). Microeconometrics: Methods and applications. New York: Cambridge University Press.

Carson, R. T., Hanemann, W. M. & Wegge, T. C. (1987). Southcentral Alaska Sport Fishing Study. Report prepared by Jones and Stokes Associates for the Alaska Department of Fish and Game. Anchorage, AK.

Carson, R. T., Hanemann, W. M. & Wegge, T. C. (2009). A nested logit model of recreational fishing demand in Alaska. Marine Resource Economics, 24, 101-129.

Cicchetti, C. J., Fisher, A. C. & Smith, V. K. (1976). An econometric evaluation of a generalized consumer surplus measure: The Mineral King controversy. Econometrica, 44, 1259-1276.

Clawson, M. & Knetsch, J. L. (1969). Economics of outdoor recreation. Baltimore, MD: Johns Hopkins University.

Creel, M. D. & Loomis, J. B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. American Journal of Agricultural Economics, 72, 434-441.

Cutter, W. B., Pendleton, L. & DeShazo, J. R. (2007). Activities in models of recreational demand. Land Economics, 83, 370-381.

Day, B. (2000). A recreational demand model of wildlife-viewing visits to the game reserves of the Kwazulu-Natal Province of South Africa (GEC-2000-08). Working paper, Centre for Social and Economic Research on the Global Environment.

Deaton, A. (1997). The analysis of household surveys: a microeconomic approach to development policy. Baltimore MD: The Johns Hopkins University Press.

Edwards, P. E. T., Parsons, G. R. & Myers, K. H. (2011). The economic value of viewing migratory shorebirds on the Delaware Bay: An application of the single site travel cost model using on-site data. Human Dimensions of Wildlife, 16, 435-444.

Egan, K. & Herriges, J. (2006). Multivariate count data regression models with individual panel data from an on-site sample. Journal of Environmental Economics and Management, 52, 567-581.

Egan, K. J., Herriges, J. A., Kling, C. L. & Downing, J. A. (2009). Valuing water quality as a function of water quality measures. American Journal of Agricultural Economics, 91, 70-86.

Englin, J., Boxall, P. & Watson, D. (1998). Modeling recreation demand in a Poisson system of equations: An analysis of the impact of international exchange rates. American Journal of Agricultural Economics, 80, 255-263.

Englin, J. & Cameron, T. A. (1996). Augmenting travel cost models with contingent behavior data. Environmental and Resource Economics, 7, 133-147.

Englin, J. & Shonkwiler, J. S. (1995). Modeling recreation demand in the presence of unobservable travel costs: Toward a travel price model. Journal of Environmental Economics and Management, 29, 368-377.

Feather, P. M. (1994). Sampling and aggregation issues in random utility model estimation. American Journal of Agricultural Economics, 76, 772-780.

Feather, P. & Hellerstein, D. (1997). Calibrating benefit function transfer to assess the conservation reserve program. American Journal of Agricultural Economics 79, 151-162.

Feather, P. & Shaw, W. D. (1999). Estimating the cost of leisure time for recreation demand models. Journal of Environmental Economics and Management, 38, 49-65.

Fisher, A. & Krutilla, J. V. (1972). Determination of optimal capacity of resource-based recreation facilities. Natural Resources Journal, 12, 417-444.

Fisher, W. L., Grambsch, A. E., Eisenhower, D. L., & Morganstein, D. R. (1991). Length of recall period and accuracy of estimates from the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. American Fisheries Society Symposium, 12, 367-374.

Freeman III, A. M. (2003). The measurement of environmental and resource values: Theory and methods (2nd ed.). Washington, DC: RFF Press.

Greene, W. H. (2007). Econometric analysis (6th ed). Upper Saddle River, NJ: Pearson Prentice Hall.

Grijalva, T. C., Berrens, R. P., Bohara, A. K., Jakus, P. M. & Shaw, W. D. (2002). Valuing the loss of rock climbing access in wilderness areas: A national-level, random-utility model. Land Economics, 78, 103-120.

Gurmu, S. & Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. Journal of Business and Economic Statistics, 14, 469-477.

Haab, T. C. & McConnell, K. E. (1996). Count data models and the problem of zeros in recreation demand analysis. American Journal of Agricultural Economics, 78, 89-102.

Haab, T. C. & McConnell, K. E. (2002). Valuing environmental and natural resources: The econometrics of non-market valuation. Cheltenham, United Kingdom: Edward Elgar.

Haener, M. K., Boxall, P. C., Adamowicz, W. L. & Kuhnke, D. H. (2004). Aggregation bias in recreation site choice models: Resolving the resolution problem. Land Economics, 80, 561-574.

Hang, D., McFadden, D., Train, K. & Wise, K. (2016). Is vehicle depreciation a component of marginal travel cost?: A literature review and empirical analysis. Journal of Transport Economics and Policy. 50(2):1-19.

Hanley, N., Bell, D. & Alvarez-Farizo, B. (2003a). Valuing the benefits of coastal water quality improvements using contingent and real behaviour. Environmental and Resource Economics, 24, 273-285.

Hanley, N., Shaw, W. D. & Wright, R. E. (2003b). The new economics of outdoor recreation. Cheltenham, United Kingdom: Edward Elgar.

Hauber, A. B. & Parsons, G. R. (2000). The effect of nesting structure specification on welfare estimation in a random utility model of recreation demand: An application to the demand for recreational fishing. American Journal of Agricultural Economics, 82, 501-514.

Hausman, J. A., Leonard, G. K. & McFadden, D. (1995). A utility-consistent, combined discrete choice and count data model assessing recreational use losses due to natural resource damage. Journal of Public Economics, 56, 1-30.

Hellerstein, D. M. (1991). Using count data models in travel cost analysis with aggregate data. American Journal of Agricultural Economics, 73, 860-866.

Hellerstein, D. M. (1992). The treatment of nonparticipants in travel cost analysis and other demand models. Water Resources Research 28, 1999-2004.

Hellerstein, D. M. & Mendelsohn, R. (1993). A theoretical foundation for count data models. American Journal of Agricultural Economics, 75, 604-611.

Herriges, J. A. & Kling, C. L. (Eds.). (1999). Valuing recreation and the environment: Revealed preference methods in theory and practice. Cheltenham, United Kingdom: Edward Elgar.

Herriges, J. A. & Kling, C. L. (Eds.). (2008). Revealed preference approaches to environmental valuation (Vol. 1). Aldershot, United Kingdom: Ashgate.

Herriges, J. A., Kling, C. L. & Phaneuf, D. J. (1999). Corner solution models of recreation demand: A comparison of competing frameworks. In J. A. Herriges & C. L. Kling (Eds.), Valuing recreation and the environment: Revealed preference methods in theory and practice. Cheltenham, United Kingdom: Edward Elgar.

Herriges, J. A. & Phaneuf, D. J. (2002). Inducing patterns of correlation and substitution in repeated logit models of recreation demand. American Journal of Agricultural Economics, 84, 1076-1090.

Hicks, R. L. & Strand, I. E. (2000). The extent of information: Its relevance for random utility models. Land Economics, 76, 374-385.

Hindsley, P., Landry, C. E. & Gentner, B. (2011). Addressing onsite sampling in recreation site choice models. Journal of Environmental Economics and Management, 62, 95-110.

Hoehn, J. P., Tomasi, T., Lupi, F. & Chen, H. Z. (1996). An economic model for valuing recreational angling resources in Michigan. Report presented to the Michigan Department of Environmental Quality. East Lansing: Michigan State University.

Hof, J. G. & King, D. A. (1982). On the necessity of simultaneous recreation demand equation estimation. Land Economics, 58, 547-552.

Hotelling, H. (1949). An economic study of the monetary valuation of recreation in the National Parks. Washington, DC: U.S. Department of the Interior, National ParkService and Recreation Planning Division.

Hunt, L. M., Boxall, P. C. & Boots, B. N. (2007). Accommodating complex substitution patterns in a random utility model of recreational fishing. Marine Resource Economics, 22, 155-172.

Hynes, S., Hanley, N. & Garvey, E. (2007). Up the proverbial creek without a paddle: Accounting for variable participant skill levels in recreational demand modelling. Environmental and Resource Economics, 36, 413-426.

Jakus, P. M., Bergstrom, J. C., Phillips, M. & Maloney, K. (2011). Modeling behavioral response to changes in reservoir operations in the Tennessee Valley region. In J. Whitehead, T. Haab & J.-C. Huang (Eds.), Preference data for environmental valuation: Combining revealed and stated approaches (pp. 253-272). New York: Routledge.

Joen, Y., Herriges, J. A., Kling, C. L. & Downing, J. (2011). The role of water quality perceptions in modelling lake recreation demand. In J. Bennett (Ed.), The international handbook on non-market environmental valuation (pp. 74-101). Cheltenham, United Kingdom: Edward Elgar.

Kaoru, Y. (1995). Measuring marine recreation benefits of water quality improvements by the nested random utility model. Resource and Energy Economics, 17, 119-136.

Kinnell, J. C., Bingham, M. F., Mohamed, A. F., Desvousges, W. H., Kiler, T. B., Hastings, E. K. & Kuhns, K. T. (2006). Estimating site choice decisions for urban recreators. Land Economics, 82, 257-272.

Laitila, T. (1999). Estimation of combined site-choice and trip-frequency models of recreational demand using choice-based and on-site samples. Economic Letters, 64, 17-23.

Landry, C. E. & Liu, H. (2009). A semi-parametric estimator for revealed and stated preference data—An application to recreational beach visitation. Journal of Environmental Economics and Management, 57, 205-218.

Landry, C. E. & Liu, H. (2011). Econometric models for joint estimation of revealed and stated preference site-frequency recreation demand models. In J. Whitehead, T. Haab & J.-C. Huang (Eds.), Preference data for environmental valuation: Combining revealed and stated approaches (pp. 87-100). New York: Routledge.

Leggett, C. G., Scherer, N., Haab, T. C., Bailey, R., Landrum, J. P. & Domanski, A. (2015). Assessing the economic benefits of reductions in marine debris at Southern California beaches: a random utility travel cost model. Industrial Economics Inc. Manuscript.

Lew, D. K., NOAA Fisheries & Larson, D. M. (2008). Valuing a beach day with a repeated nested logit model of participation, site choice, and stochastic time value. Marine Resource Economics, 23 (3), 233-252.

Lupi, F. & Feather, P. M. (1998). Using partial site aggregation to reduce bias in random utility travel cost models. Water Resources Research, 34, 3595-3603.

Lupi, F., Hoehn, J. P., and Christie, G. C. (2003). Using an economic model of recreational fishing to evaluate the benefits of sea lamprey (*Petromyzon marinus*) control on the St. Marys River. Journal of Great Lakes Research 29 (Supplement 1), 742-754.

Martinez-Espiñeira, R. & Amoako-Tuffour, J. (2008). Recreation demand analysis under truncation, overdispersion, and endogenous stratification: An application to Gros Morne National Park. Journal of Environmental Management, 88, 1320-1332.

Massey, D. M., Newbold, S. C. & Gentner, B. (2006). Valuing water quality changes using a bioeconomic model of a coastal recreational fishery. Journal of Environmental Economics and Management, 52 (1), 482-500.

McConnell, K. E. (1977). Congestion and willingness to pay: A study of beach use. Land Economics, 53, 185-195.

McConnell, K. E. (1992). On-site time in the demand for recreation. American Journal of Agricultural Economics, 74, 918-925.

McConnell, K. E. & Strand, I. (1981). Measuring the cost of time in recreation demand analysis: An application to sportfishing. American Journal of Agricultural Economics, 63, 153-156.

McConnell, K. E. & Strand, I. E. (1994). The economic value of mid and south Atlantic sportfishing. Report presented to the University of Maryland, U. S. Environmental Protection Agency, U.S. National Marine Fisheries Service, and the National Oceanic and Atmospheric Administration. College Park: University of Maryland.

McConnell, K. E., Strand Jr., I. E. & Blake-Hedges, L. (1995). Random utility models of recreational fishing: Catching fish using a Poisson process. Marine Resource Economics, 10, 247-261.

McConnell, K. E., Strand, I. E. & Bockstael, N. E. (1990). Habit formation and the demand for recreation: Issues and a case study. In V. K. Smith, A. D. Witte & A. N. Link (Eds.), Advances in applied microeconomics 5 (pp. 217-235). New York: Jai Press.

McFadden, D. (2001). Economic choices. American Economic Review, 91, 351-378.

Mendelsohn, R., Hof, J., Peterson, G. & Johnson, R. (1992). Measuring recreation values with multiple destination trips. American Journal of Agricultural Economics, 74, 926-933.

Moeltner, K. (2003). Addressing aggregation bias in zonal recreation models. Journal of Environmental Economics and Management, 45, 128-144.

Moeltner, K. & Englin, J. (2004). Choice behavior under time-variant quality: State dependence versus 'play-it-by-ear' in selecting ski resorts. Journal of Business and Economic Statistics, 22, 214-224.

Moeltner, K. & Shonkwiler, J. S. (2005). Correcting for on-site sampling in random utility models. American Journal of Agricultural Economics, 87, 327-339.

Morey, E. R. (1981). The demand for site-specific recreational activities: A characteristics approach. Journal of Environmental Economics and Management, 8, 345-371.

Morey, E. R. (1999). Two RUMS uncloaked: Nested-logit models of site choice and nested-logit models of participation and site choice. In J. A. Herriges & C. L. Kling (Eds.), Valuing recreation and the environment: Revealed preference methods in theory and practice (pp. 65-120). Cheltenham, United Kingdom: Edward Elgar.

Morey, E. R., Shaw, W. D. & Rowe, R. D. (1991). A discrete-choice model of recreational participation, site choice, and activity valuation when complete trip data are not available. Journal of Environmental Economics and Management, 20, 181-201.

Morey, E. R. & Waldman, D. M. (1998). Measurement error in recreation demand models: The joint estimation of participation, site choice, and site characteristics. Journal of Environmental Economics and Management, 35, 262-276.

Morey, E. R. & Waldman, D. M. (2000). Joint estimation of catch and other travel-cost parameters —Some further thoughts. Journal of Environmental Economics and Management, 40, 82-85.

Morgan, O. A. & Huth, W. L. (2011). Using revealed and stated preference data to estimate the scope and access benefits associated with cave diving. Resource and Energy Economics, 33, 107-118.

Murdock, J. (2006). Handling unobserved site characteristics in random utility models of recreation demand. Journal of Environmental Economics and Management, 51, 1-25.

Murray, C., Sohngen, B. & Pendelton, L. (2001). Valuing water quality advisories and beach amenities in the Great Lakes. Water Resources Research, 37, 2583-2590.

Offenbach, L. A. & Goodwin, B. K. (1994). A travel-cost analysis of the demand for hunting trips in Kansas. Review of Agricultural Economics, 16, 55-61.

Parsons, G. R. (1991). A note on choice of residential location in travel cost demand models. Land Economics, 67, 360-364.

Parsons, G. R. (2013). The travel cost model. In J. Shogren (Ed.), Encyclopedia of energy, natural resource, and environmental economics (Vol. 3: Environment, pp. 349-358). Amsterdam: Elsevier.

Parsons, G. R., Chen, Z., Hidrue, M., Lilley, J., Standing, N. & Lilley, J. (2013). Valuing beach width for recreational use: Combining revealed and stated preference data. Marine Resource Economics, 28, 221-241.

Parsons, G. R., Jakus, P. M. & Tomasi, T. (1999a). A comparison of welfare estimates from four models for linking seasonal recreational trips to multinomial logit models of site choice. Journal of Environmental Economics and Management, 38, 143-157.

Parsons, G. R., Kang, A. K., Leggett, C. G. & Boyle, K. J. (2009). Valuing beach closures on the Padre Island National Seashore. Marine Resource Economics, 24, 213-235.

Parsons, G. R. & Kealy, M. J. (1992). Randomly drawn opportunity sets in a random utility model of lake recreation. Land Economics, 68, 93-106.

Parsons, G. R. & Massey, D. M. (2003). A random utility model of beach recreation. In N. Hanley, W. D. Shaw & R. E. Wright (Eds.), The new economics of outdoor recreation (pp. 241-267). Cheltenham, United Kingdom: Edward Elgar.

Parsons, G. R., Massey, D. M. & Tomasi, T. (1999b). Familiar and favorite sites in a random utility model of beach recreation. Marine Resource Economics, 14, 299-315.

Parsons, G. R. & Needelman, M. S. (1992). Site aggregation in a random utility model of recreation. Land Economics, 68, 418-433.

Parsons, G. R., Plantinga, A. J. & Boyle, K. J. (2000). Narrow choice sets in a random utility model of recreation demand. Land Economics, 76, 86-99.

Parsons, G. R. & Stefanova, S. (2011). Gauging the value of short-time site closures in a travel-cost random utility model of recreation demand with a little help from stated preference data. In J. Whitehead, T. Haab & J.-C. Huang (Eds.), Preference data for environmental valuation: Combining revealed and stated approaches (pp. 239-252). New York: Routledge.

Parsons, G. R. & Wilson, A. J. (1997). Incidental and joint consumption in recreation demand. Agricultural and Resource Economics Review, 26, 1-6.

Peters, T., Adamowicz, W. L. & Boxall, P. C. (1995). Influence of choice set considerations in modeling the benefits from improved water quality. Water Resources Research, 31, 1781-1787.

Phaneuf, D. J. (2002). A random utility model for total maximum daily loads: Estimating the benefits of watershed-based ambient water quality improvements. Water Resources Research, 38, 1254-1264.

Phaneuf, D. J., Kling, C. L. & Herriges, J. A. (2000). Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand. Review of Economics and Statistics, 82, 83-92.

Phaneuf, D. J. & Siderelis, C. (2003). An application of the Kuhn-Tucker model to the demand for water trail trips in North Carolina. Marine Resource Economics, 18 (1), 1-14.

Phaneuf, D. J. & Smith, V. K. (2005). Recreation demand models. In K.-G. Mäler & J. R. Vincent (Eds.), Handbook of environmental economics (pp. 671-761). Amsterdam: Elsevier.

Provencher, B. & Bishop, R. C. (1997). An estimable dynamic model of recreation behavior with an application to Great Lakes angling. Journal of Environmental Economics and Management, 33, 107-127.

Provencher, B. & Moore, R. (2006). A discussion of "using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model." Environmental and Resource Economics, 34, 117-124.

Riera, P., McConnell, K. E., Giergiczny, M. & Mahieu, P. (2011). Applying the travel cost method to Minorca beaches: Some policy results. In J. Bennett (Ed.), The international handbook on non-market environmental valuation (pp. 60-73). Cheltenham, United Kingdom: Edward Elgar.

Scarpa, R., Thiene, M. & Train, K. (2008). Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps. American Journal of Agricultural Economics, 90, 994-1010.
Shaw, D. (1988). On-site samples' regression: Problems of non-negative integers, truncation, and endogenous stratification. Journal of Econometrics, 37, 211-223.
Shaw, W. D. & Jakus, P. M. (1996). Travel cost models of the demand for rock climbing. Agricultural and Resource Economics Review, 25, 133-142.
Shaw, W. D. & Ozog, M. T. (1999). Modeling overnight recreation trip choice: Application of a repeated nested multinomial logit model. Environmental and Resource Economics, 13, 397-414.
Shonkwiler, J. S. & Shaw, W. D. (1996). Hurdle count-data models in recreation demand analysis. Journal of Agricultural and Resource Economics, 21(2), 210-219.
Siderelis, C., Brothers, G. & Rea, P. (1995). A boating choice model for the valuation of lake access. Journal of Leisure Research, 27, 264-282.
Small, K. A. & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. Econometrica, 49, 105-130.
Smith, V. K. & Desvousges, W. H. (1985). The generalized travel cost model and water quality benefits: A reconsideration. Southern Economics Journal, 52, 371-381.
Smith, V. K., Desvousges, W. H. & McGivney, M. P. (1983). The opportunity cost of travel time in recreation demand models. Land Economics, 59, 259-278.
Sohngen, B. (2000). The value of day trips to Lake Erie beaches. Unpublished report, Department of Agricultural, Environmental and Development Economics, Ohio State University, Columbus.
Stefanova, S. (2009). Measuring the recreational value of changes in beach access, beach width, and vehicle access in the Mid-Atlantic region: Application of random utility models. (Doctoral dissertation). University of Delaware, Newark.
Swait, J., Adamowicz, W. & van Bueren, M. (2004). Choice and temporal welfare impacts: Incorporating history into discrete choice models. Journal of Environmental Economics and Management, 47, 94-116.
Tay, R., McCarthy, P. S. & Fletcher, J. J. (1996). A portfolio choice model of the demand for recreational trips. Transportation Research Part B: Methodological, 30, 325-337.
Thiene, M. & Scarpa, R. (2009). Deriving and testing efficient estimates of WTP distributions in destination choice models. Environmental and Resource Economics, 44, 379-395.
Timmins, C. & Murdock, J. (2007). A revealed preference approach to the measurement of congestion in travel cost models. Journal of Environmental Economics and Management, 53, 230-249.
Train, K. (1986). Qualitative choice analysis: Theory, econometrics, and an application to automobile demand. London: MIT Press.
Train, K. E. (1998). Recreation demand models with taste differences over people. Land Economics, 74, 230-239.
Train, K. (2009). Discrete choice methods with simulation (2nd ed.). New York: Cambridge University Press.
Train, K., McFadden, D. & Johnson, R. (2000). Discussion of Morey and Waldman's "Measurement error in recreation demand models." Journal of Environmental Economics and Management, 40, 76-81.
Trice, A. H. & Wood, S. E. (1958). Measurement of recreation benefits. Land Economics, 34, 195-207.
Von Haefen, R. H. (2008). Latent consideration sets and continuous demand systems. Environmental and Resource Economics, 41, 363-379.
Whitehead, J. C., Dumas, C. F., Herstine, J., Hill, J. & Buerger, R. (2008). Valuing beach access and width with revealed and stated preference data. Marine Resource Economics, 23, 119-135.
Whitehead, J. C., Haab, T. C. & Huang, J.-C. (2000). Measuring recreation benefits of quality improvements with revealed and stated behavior data. Resource and Energy Economics, 22, 339-354.

Whitehead, J., Haab, T. & Huang, J.-C. (Eds.). (2011). Preference data for environmental valuation: Combining revealed and stated approaches. New York: Routledge.

Yeh, C.-Y., Haab, T. C. & Sohngen, B. L. (2006). Modeling multiple-objective recreation trips with choices over trip duration and alternative sites. Environmental and Resource Economics, 34, 189-209.

Whitehead, J. C,, Phaneuf, D. J., Dumas, C. F., Herstine, J., Hill, J. & Buerger, B. (2010). Convergent validity of revealed and stated recreation behavior with quality change: a comparison of muliple and single site demands. Environmental and Resource Economics, 45, 91-112.

Chapter 7
Hedonics

Laura O. Taylor

Abstract This chapter covers the current theory and empirical methods in hedonic valuation of environmental and natural resources. A framework is first presented that links hedonic price functions to theoretically correct welfare measures for changes in environmental amenities. The major empirical methods for estimating a hedonic price function are discussed beginning with data construction and basic estimation approaches, and progressing through to techniques for addressing endogenous regressors including spatial econometrics and quasi-experimental methods. The use of the hedonic price function for obtaining measures of welfare change for changes in environmental amenities are also presented. Sorting models and second-stage demand analysis in both a single-market and multiple-market context are described. Applications and examples from housing and labor markets are used throughout to illustrate concepts covered.

Keywords Hedonic method · Implicit prices · First-stage estimation · Second-stage estimation · Quasi-experimental methods · Sorting models · Welfare measures

Heterogeneous or differentiated goods are products whose characteristics vary in such a way that there are distinct product varieties even though the product is sold in one market (e.g., cars, computers, houses). The variation in product variety gives rise to variations in product prices within each market. The hedonic method relies on market transactions for these differentiated goods to determine the implied value or implicit price of characteristics. For instance, by observing the price differential between two product varieties that vary only by one characteristic (e.g., two identical cars, but one has more horsepower than the other), we indirectly observe the monetary trade-offs individuals are willing to make with respect to the changes in this characteristic, and the value of the increase in horsepower is the difference in the prices of the two cars. As such, the hedonic method is an indirect valuation

L.O. Taylor (✉)
North Carolina State University, Raleigh, USA
e-mail: laura_taylor@ncsu.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_7

method where we do not observe the value consumers have for the characteristic directly, but infer it from observable market transactions.

The most common application of hedonic theory in environmental valuation involves housing markets. The choice of housing location and, therefore, neighborhood amenities, is observable. Often, location choice is directly linked to an environmental amenity of interest. For example, housing locations can offer different scenic vistas (Paterson and Boyle 2002; Bin et al. 2008), or they can impose greater perceived risks by placing a household closer to perceived hazards, such as shale gas extraction activities (Gopalakrishnan and Klaiber 2014). As such, the choice of a house and its associated price implies an implicit choice over the environmental amenities (or disamenities) proximate to the house and their implicit prices.

To see how implicit prices for environmental goods are revealed through market transactions, imagine the following scenario in which there are two identical lakes, each with 100 identical homes surrounding them. All homes are lakefront, and all the characteristics of the homes themselves, the land, and the neighborhoods are identical across the properties. At the current equilibrium price of $200,000 per house, all 200 homes are equally preferred. Now, imagine that water clarity at one lake, Lake A for example, is improved. We assume that the improved water clarity is preferred by all households. Now if any home on Lake A were offered at the original equilibrium price of $200,000, consumers would uniformly prefer this house to any house on Lake B. In other words, at the current price, there would be excess demand for the houses located on Lake A, and as such, the price of these houses must rise to bring the market into equilibrium. The price differential that results from the change in water clarity at Lake A is the implicit price consumers are willing to pay for that incremental increase in water clarity. This willingness to pay for water clarity is indirectly revealed to us through the market prices of the homes. For instance, if in the new equilibrium, houses on Lake A sell for $210,000 while houses on Lake B sell for $200,000, the implicit price associated with the increased water clarity is $10,000.

Of course, housing markets aren't so simple: housing choice depends on many characteristics, such as structure of the house, amenities of the land, neighborhood, and location. Yet, the fundamental intuition behind the hedonic method extends easily. By observing the choices consumers make over heterogeneous commodities with varying prices, we can estimate the implicit prices of component characteristics. These implicit prices or hedonic prices, under certain conditions, are equal to Marshallian willingness to pay (WTP) or allow us to recover WTP.

Hedonic analyses have been reported as early as Waugh's (1928) analysis of quality factors influencing asparagus pricing and have been applied to markets as varied as automobiles, computers, VCRs, household appliances, art, and agricultural commodities such as organic produce and wine. This chapter focuses primarily on housing markets and how they can be used to value environmental amenities. The application of the hedonic method to labor markets is also briefly reviewed in Sect. 7.5 (Bockstael and McConnell, 2007, provided a detailed review). Other reviews of the hedonic method can be found in Palmquist (2006), and Phaneuf and Requate (in press).

Hedonic analysis of markets for differentiated goods consists of two related steps often referred to as a first-stage and second-stage analysis. In a first-stage analysis, the hedonic price function is estimated using information about the sales prices of a differentiated product and the characteristics of the product. This analysis allows researchers to recover the implicit prices of characteristics and reveals information on the underlying preferences for these characteristics, which is discussed in Sect. 7.3. First-stage analyses are the most common application of the hedonic method because the needed economic insights often require only marginal price information, and the data are generally readily available. An introduction to the empirical method and data collection is presented in Sects. 7.1 and 7.2.2.

Although econometric estimation of the hedonic price function is conceptually straightforward, housing and other real estate markets have unique challenges. In particular, hedonic housing applications are usually interested in environmental features that vary over space, such as air quality across an urban area. Concerns about unobservable characteristics of housing and their neighborhoods that co-vary over space with environmental features of interest have led to an increased use of spatial econometrics and quasi-experimental methods to address the potential for biased coefficient estimates due to endogenous regressors. Both of these approaches are discussed in Sect. 7.2.3.

Once a first-stage analysis is completed, researchers can combine information on the implicit prices obtained in the first-stage with data on household characteristics to estimate demand functions for the characteristics or utility parameters. This step is referred to as a second-stage analysis and is discussed in Sect. 7.4. While second-stage analysis had traditionally been less common due to data demands, this is changing rapidly as data acquisition costs have decreased. Second-stage analyses are also very important because environmental policy analyses frequently require welfare analyses of large-scale environmental changes, such as air quality changes across an entire metropolitan area, and these situations require that underlying demands or utility parameters be recovered.

Before discussing the implementation of first- and second-stage analyses, the next section reviews the theory that provides a framework for understanding the market process generating a hedonic equilibrium. For the purposes of nonmarket valuation, Rosen's (1974) seminal article provided this theory, and it is important because it developed the utility theoretic framework that establishes the connections between consumers' preferences for characteristics of heterogeneous goods and the equilibrium price function.

7.1 The Hedonic Price Function: Theory

This chapter discusses the hedonic theory using the following notation and assumptions. Let Z represent the differentiated product with characteristics $\underline{z} = z_1, z_2, z_3, \ldots z_n$. The differentiated product is assumed to be sold in a perfectly competitive market, and the interactions of the many producers and consumers together determine an equilibrium price schedule for the differentiated product, $P(\underline{z})$. The equilibrium price for any one particular variety of a differentiated good (e.g., a specific house) is a function of the characteristics of that particular variety. As such, the consumer can determine the price he or she pays for the good by choosing which model to purchase. However, it is important to note that the consumer takes the entire price schedule $P(\underline{z})$ as exogenous.

Consumer utility is defined over two goods: Z, the differentiated good, and x, a composite product representing all other goods (i.e., income left over after purchasing Z). Consumer j, with demographic characteristics α^j, has utility defined as

$$U^j\left(x, z_1, z_2, \ldots, z_n; \alpha^j\right). \tag{7.1}$$

If one assumes that the consumer purchases only one unit of the differentiated product, a reasonable assumption for the housing market, the budget constraint is $y^j = x + P(\underline{z})$. The consumer seeks to maximize utility by choosing the model of the differentiated product, $\underline{z}$, and the amount of x to purchase, subject to his or her budget constraint. The consumer will choose $\underline{z}$ and x such that the following is satisfied for each z_i:

$$\frac{\partial P}{\partial z_i} = \frac{\partial U/\partial z_i}{\partial U/\partial x}, \tag{7.2}$$

which indicates that the marginal rate of substitution between any characteristic, z_i, and the composite numeraire good, x, is equal to the rate at which the consumer can trade z_i for x in the market (i.e., the ratio of the implicit price for z_i and the price of the numeraire, which is, by definition, equal to one).

A second convenient way to describe the optimization process is to define the optimal bid a consumer will make for any specific product variety. The bid function, θ, describes the relationship between the dollar bid Consumer j will make for Z as one or more of its component characteristics are changed while utility and income remain constant. Equation (7.1) can be used to formally define the bid function by recognizing that income less the bid a consumer makes for Z is the amount of money left over to spend on the numeraire, x. Thus, the relationship

$$U^j\left(y_0 - \theta, \underline{z}, \alpha^j\right) \equiv U_0^j, \tag{7.3}$$

indicates how a consumer's optimal bid must vary in response to changes in z if utility and income are held constant, where y is exogenous income, and U_0 is a fixed level of utility. Solving Eq. (7.3) for θ indicates that $\theta^j = \theta(\underline{z}, y_0, U_0^j, \alpha^j)$. Maximizing utility in (7.3) yields the result that the marginal bid a consumer is willing to make for z_i ($\partial\theta/\partial z_i$, which is denoted as θ_i), will equal the marginal rate of substitution between any characteristic, z_i, and x. Relating this result to Eq. (7.2) indicates that the necessary condition for utility maximization is that the marginal bid a consumer places for a characteristic must equal the marginal price of that characteristic ($\partial P(\underline{z})/\partial z_i$, which is denoted as P_i).

For the supply side of the market, one can describe a firm with characteristics δ^k as seeking to maximize profits: $\Pi = H * P(\underline{z}) - C(H, \underline{z}, \delta^k)$, where H is the number of units of Z that the firm produces, and $C(\cdot)$ is a well-behaved cost function. Again, the firm faces the exogenous equilibrium price schedule, $P(\underline{z})$, when determining its

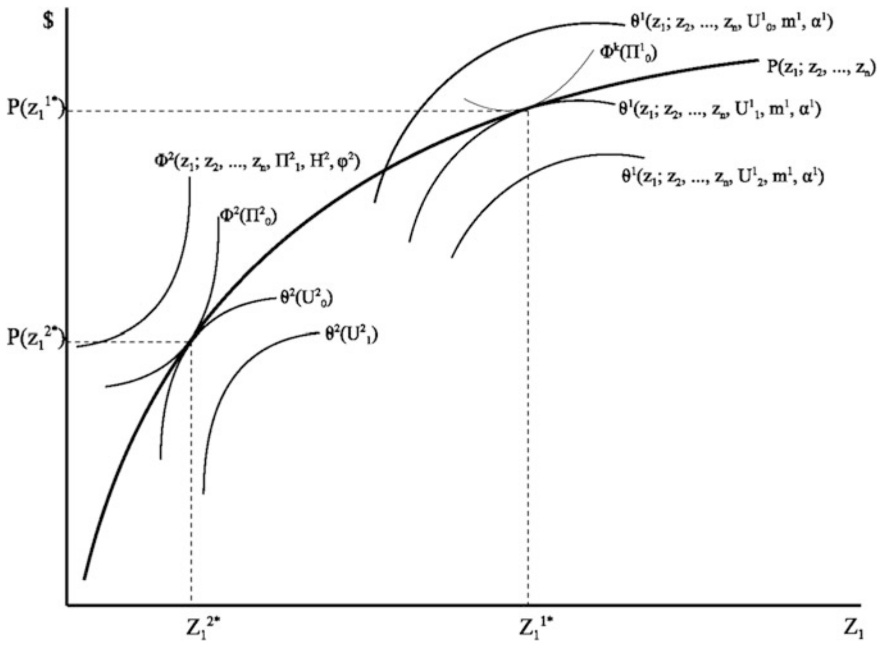


Fig. 7.1 The hedonic price function

choices. Although firms can affect the prices they receive for their products by varying the characteristics of the product, no single firm can affect the price schedule. In this formulation, we assume the firm produces only one model of Z. Thus, the firm chooses what type to produce, Z^k, and then chooses how many of that type to produce.[1] Similar to the consumer's problem, one can describe the behavior of a firm in this differentiated goods market by an offer function, $\varphi^k = \varphi(\underline{z}; H, \Pi_0, \delta^k)$, which describes the amount of money a firm is willing to accept for any particular variety of Z, holding constant the number of units of that variety the firm produces, H, and its level of profit, Π_0. The offer function is defined by $\Pi_0 = H * \varphi^k - C(H, \underline{z}, \delta^k)$, and at the optimum, it will be the case that the marginal price a firm is willing to accept for $\underline{z}_i$, φ_{zi}, will equal the marginal cost of producing that characteristic per unit of the differentiated good, C_{zi}/H.

The bid and offer functions and their properties may be easily described using Fig. 7.1. This figure illustrates the equilibrium price schedule, $P(\underline{z})$ as it varies with changes in z_1, holding the level of all other characteristics constant. $P(\underline{z})$ is drawn such that the total price paid for z_1 increases at a decreasing rate, which one might expect in many applications. For instance, say z_1 represents the number of square

[1]It may also be assumed that firms produce multiple types of the good and that the cost function is separable in each type of product. The maximization problem for the firm choosing an entire product line in this case is equivalent to choosing each product type separately, as described here.

feet of living space in a house. We might expect a smaller price differential between a 5,000- and a 5,300-square-foot house as compared to the price differential between a 1,000- and a 1,300-square-foot house.

Also depicted in Fig. 7.1 are the bid functions for two consumers, θ^1 and θ^2. Along any bid function contour, only the level of z_1 changes; the level of all other characteristics, income, and utility are constant. Bid functions are assumed to be concave in $\underset{\sim}{z}$ (i.e., optimal bids increase at a decreasing rate in $\underset{\sim}{z}$), and higher levels of utility for a consumer are represented by bid function contours closer to the horizontal axis. Intuitively, a lower bid for the same level of z_1 implies a higher level of utility because more money is left over to spend on x.[2] The optimal choice of z_1 is where the consumer reaches the lowest possible bid function while still being able to participate in the market. For Consumer 1, this occurs at a quantity z_1^{1*} and a total bid price of $P(z_1^{1*})$, which is the point at which the bid function is tangent to the equilibrium price schedule in Fig. 7.1. For Consumer 2, this optimal choice occurs at z_1^{2*}, and the consumer pays a total price of $\mathrm{P}(z_1^{2*})$. At each consumer's optimal choice of z_1, the marginal bid ($\partial\theta/\partial z_1$) equals the marginal price ($\partial P(\underset{\sim}{z})/\partial z_1$).

Offer functions for two firms, φ^1 and φ^2, are also depicted in Fig. 7.1. As indicated in the figure, these functions are convex in $\underset{\sim}{z}$ as optimal offers increase at an increasing rate, holding profits and all else constant. Offer functions farther away from the horizontal axis represent higher levels of total profit. The firm's optimal choice of what level of z_1 to produce in each of its H units of the differentiated product is where the firm reaches the highest possible offer function while still being able to participate in the market. For Firm 2, this occurs at quantity z_1^{2*} and a total offer price of $P(z_1^{2*})$.

Figure 7.1 illustrates that the hedonic price function is simply an envelope of the equilibrium interactions between all buyers and sellers of a differentiated good. As such, it is possible for the hedonic price function to take any shape. The key point is that with relatively small data requirements, such as information on product types and their sales prices, we can recover the marginal implicit prices for any component characteristic of Z. The marginal price is equal to the revealed marginal WTP for that characteristic by consumers, $\partial\theta/\partial z_i$. In the Maine lakes example, Eq. (7.1) indicates that consumers' value for a one-unit (1 foot) increase in a property's footage of lake frontage is \$83 ($\partial\theta/\partial\mathrm{FRONT} = \partial P/\partial\mathrm{FRONT}$). This is the fundamental insight underlying the hedonic method. Researchers seek to estimate the parameters of the hedonic price function so they can recover information about the marginal value consumers place on characteristics.

[2]Another property of the bid function is that optimal bids increase proportionally with income $\partial\theta/\partial y = 1$. If income increases by \$1, the consumer's optimal bid must increase by the same amount to hold utility constant.

7.2 The Hedonic Price Function: Estimation

This section discusses the details of estimating a hedonic price function as well as some important threats to validity that researchers should consider as they attempt to uncover unbiased marginal value estimates. Table 7.1 summarizes these steps and considerations. As indicated, the first step is to gather the appropriate data; the basics of this task are discussed in Sect. 7.2.1. Common econometric specifications for the hedonic price function are discussed next in Sect. 7.2.1.1, which is necessary to ground the discussion of sample frame, data quality, and econometric methods that address data shortcomings that are presented in the remainder of Sect. 7.2. Before discussing each of the steps for estimating a hedonic price function, this section begins with a concrete example of a hedonic price function for housing that is based on a study by Boyle et al. (1999). Some of the study's results that were not reported in the journal article are also included in this chapter.

Boyle et al. (1999) estimated hedonic price functions for lakefront properties in Maine. Sales prices of properties, mostly summer cottages, were analyzed as a function of the characteristics of the structure on the property and important characteristics of the land, such as its location relative to the nearest town and the water quality of the lake on which the property is located. Lakes in Maine are known for their high water quality; however, this quality is being compromised in many Maine lakes by eutrophication resulting from nonpoint pollution. The physical manifestation of eutrophication is reduced water clarity. Thus, water clarity is the measure of lake water quality that is used by the authors in the hedonic price function.

Boyle et al. (1999) estimated a hedonic price function for each of four geographically distinct markets. The estimated hedonic price function, $P(\underline{z})$, for one market is:

$$\begin{aligned} P = {} & 25{,}899 + 6{,}790 \cdot \ln(\mathrm{SQFT}) + 83 \times \mathrm{FRONT} - 3{,}919 \times \mathrm{DNSTY} + 1{,}516 \times \mathrm{DIST} \\ & + 11{,}572 \times \mathrm{HEAT} + 23{,}465 \times \mathrm{BATH} - 17{,}579 \times \mathrm{LKWATER} + 2.057 \times \mathrm{WQ}, \end{aligned} \tag{7.4}$$

where the dependent variable is the sales price of a property, and the independent variables are, respectively, square feet of the structure on the property (mean = 750 square feet), the property's frontage on the lake (mean = 143 feet), the number of lots per 1,000 feet of frontage adjacent to the property (mean = 8.9 lots), the distance from the property to the nearest town (mean = 9.4 miles), and dummy variables that designate whether or not the structure has central heating, a full bath, or if the property uses lake water as its primary source of water. The last independent variable in Eq. (7.4) is a measure of water quality (WQ), which is equal to the area of the lake on which the property is located (mean = 4,756 a) multiplied by the depth of water clarity for the lake (mean = 3.9 m).

Table 7.1 Summary of steps for estimating the hedonic price function

Step 1	Collect data on property values and characteristics (Sects. 7.2.1 and 7.2.2)
	• Sales price: preferred measure of value, may need to consider selection bias • Pay careful attention environmental amenity/disamenity measurement • Develop appropriate neighborhood and locational variables • Develop a geographic information systems (GIS) database linked to census and environmental data • Consider appropriate sample frame for data (both across time and space)
Step 2	Choose functional form for the hedonic price function (Sect. 7.2.1.1)
	• Simple linear function in price and all characteristics generally not appropriate • Semilog functional form often used. Newer research suggests more flexible functional forms combined with spatial controls outperform simpler semilog form • Researcher judgment must be applied, and expectations about relationships between certain characteristics and sales price will guide choice of functional form
Step 3	Consider endogenous regressors potential due to omitted variables (Sect. 7.2.3)
	• Omitted variable concerns typically centered on omitted spatially varying characteristics • Spatial error model appropriate if omitted variables are thought to be independent of regressors • Spatial lag model and quasi-experimental designs can alleviate omitted variable bias
Step 4	Compute welfare measures (Sect. 7.3)
	• For marginal or nonmarginal changes in an amenity that are localized within a part of the market, the change in sales price resulting from the change in the amenity is the measure of net benefits if there are no transactions costs associated with moving between properties. If there are transactions costs, the change in price net of transactions costs measures net benefits (or is an upper bound on net benefits) • If a quasi-experimental design is used to estimate the change in sales price resulting from a nonmarginal change in an amenity, resultant capitalization rates are not likely to equal to net benefits • For nonlocalized changes in amenities, a second-stage demand analysis or sorting model approach is most appropriate for computing net-benefits

7.2.1 Data and Estimation Basics

At the most basic level, the data needs for estimating the hedonic price function and associated implicit prices are fairly simple. First, as indicated in Table 7.1, the appropriate dependent variable is sales price because the goal is to estimate the equilibrium price schedule, $P(\underline{z})$, which is a function that relates transaction prices (sales prices) to the characteristics of a product.[3] As with any durable asset, the sales price of a property represents the discounted present value (PV) of all future rents (R) from the property:

[3]Note: Alternatives to transactions prices, such as appraised value and owner-reported values, are also commonly used.

$$\mathrm{PV} = \sum_{t=1}^{T} \frac{E[R_t]}{(1+r)^t}, \tag{7.5}$$

where r is the discount rate, and T is the expected life of the property. As Eq. (7.5) makes clear, expected changes in future benefits associated with the house would be incorporated into current sales prices in a discounted fashion. Similarly, one can think of the implicit price for an individual characteristic as representing the current value of the discounted expected stream of benefits from that characteristic.

In some cases, the researcher might be interested in understanding how rental markets respond to changes in environmental quality (Grainger 2012; Baranzini and Schaerer 2011; Taylor and Smith 2000). In this case, the dependent variable is the contracted rental price and implicit prices are actually implicit rents, representing the additional value to rent from an additional unit of a particular characteristic. Rental prices are typically monthly rents but can be weekly rental prices in the case of vacation rentals. It is important to note future changes in amenities are not expected to be capitalized into current rents (as Eq. (7.5) makes clear). Although this does not diminish the usefulness of rental prices for hedonic analysis, the researcher has to be clear when interpreting implicit prices, noting that they represent the value of existing amenities in the current time period rather than changes in asset values (see also Epple et al., 2013).

Given a set of transaction prices, the next step is to identify the set of characteristics, $\underset{\sim}{z}$, that describe the good and influence its sales price. For example, the lakefront cottages used by Boyle et al. (1999) were assumed to have eight primary characteristics that determined price (as shown in Eq. (7.4)). Researchers must use their knowledge of the market to determine what characteristics are relevant for determining price in that market. However, there are important rules of thumb that apply to all circumstances. First, only those product characteristics that affect sales prices in a meaningful way are to be included as regressors. For example, the color of a product type often does not affect its price. Second, characteristics of the buyer and seller of the product *do not* belong in the hedonic price regression when markets are competitive (see also Taylor, 2008). This is not to say that characteristics of neighbors surrounding a home cannot be included—they can and should be included because they describe the neighborhood in which a home is located.

Three categories of regressors that are common to all hedonic housing studies are those that describe (1) characteristics of the house and its lot, (2) features of the home's neighborhood, and (3) the property's locational characteristics. There is no easy answer to the question of exactly what variables must be included within these three categories, but each is typically represented. For instance, housing features typically include at least the structure's size as measured by square footage or the number of bedrooms and baths. The age of the structure is usually included. Quality indicators, such as the presence of a fireplace, are sometimes included when available. The lot size is an important characteristic that is usually available and included in the analysis. Characteristics of lots are sometimes included, such as the amount of tree cover (Sander et al. 2010) or the view of mountains or water bodies afforded by the property (Bin et al. 2008; Baranzini and Schaerer 2011). Equation (7.4) indicates that Boyle et al.'s (1999) hedonic price function for summer vacation homes in Maine included square footage of the home, whether or not the summer home had central heat and a full bath, and whether it had lake water as its source of indoor water.

Neighborhood characteristics describe the immediate surroundings of a home and often include measures of the quality of the schools to which the home is associated and information about the general characteristics of the home's neighbors, such as median household income or racial composition. Neighborhood characteristics may also include an environmental amenity, such as ambient air quality, or undeveloped open space as a percentage of total land area in the neighborhood as a whole. In the Maine summer home example (Eq. (7.4)), the water quality of the lake on which a home was located and the density of development surrounding the home in question represented the key neighborhood characteristics.

Locational characteristics are conceptually similar to neighborhood characteristics in that they focus on a home's spatial relationship to other features of interest. In theory, locational characteristics vary by house, such as the distance of a house to an airport. In many cases, researchers will have access to very precise information on where a home is located and will be able to compute the exact distance of a home to features of interest. Boyle et al. (1999) included the distance of each summer home to the nearest town, and this varied by home (see Eq. (7.4)). However, in some cases the researcher may only know the neighborhood in which a home is located. Measuring locational characteristics at a coarser level, such as proximity of a neighborhood's center point to an airport, may be sufficient to capture the meaningful variation of the regressor. In this case, the distinction between a neighborhood feature and a locational feature is purely semantic.

There are many locational features that have been the focal point of hedonic studies using housing markets. Recent applications include studies that have focused on a home's proximity to amenities such as parks (Liu et al. 2013), oceans and water bodies (Landry and Hindsley 2011), and other forms of general open space (see McConnell and Walls, 2005, and Brander and Koetse, 2011, for recent reviews). The value of proximity to urban amenities, such as central business districts, transit hubs, or shopping centers has also been explored (e.g., Hess and Almeida 2007; Matthews and Turnbull 2007). Proximity to disamenities that have been studied include proximity to environmentally contaminated properties (e.g., Taylor et al. 2016; Zabel and Guignet 2012; Gamper-Rabindran and Timmins 2013; Kiel and Williams 2007), landfills (Hite et al. 2001), hog farms (Kim and Goldsmith 2009), power plants (Davis 2011), nuclear waste shipment routes (Gawande et al. 2013), and wind farms (Heintzelman and Tuttle 2012), among many others. The previous examples describe proximity of a property to a feature of interest, which conveys some form of exposure to the amenity or disamenity. Ambient conditions of a property have also been studied extensively, such as air quality (e.g., Grainger 2012; Chay and Greenstone 2005; Smith and Huang 1995) and noise reductions (e.g., Baranzini and Ramirez 2005; Pope 2008; Boes and Nüesch 2011).

There are a number of practical questions that an empirical investigator faces. What variables within each category described should be included in the hedonic equation? Should as many variables as possible be included for each category within the hedonic equation? These questions must be thoughtfully addressed. In an empirical investigation, the researcher should gather information from knowledgeable individuals (e.g., Realtors), review related hedonic studies, and test the robustness of his or her results to assumptions made during the modeling exercise. To help guide the researcher in his or her thinking, however, the goal should be remembered: an unbiased estimate of the implicit price of a particular characteristic. While inclusion of extraneous regressors may result in multicollinearity issues, this is a question of efficiency, not bias. The biggest threat to estimation of unbiased implicit prices is omitted variables. The potential for omitted variable bias is especially important in housing studies that focus on a locational characteristic given the multitude of locational features that may and that may or may not be known to the researcher. Given the importance of locational variables in nonmarket valuation, their measurement and how inclusion or omission of spatial features influences the researcher's ability to identify the implicit prices of any one particular locational feature of interest are discussed in detail in Sects. 7.2.2.2 and 7.2.3, respectively.

7.2.1.1 Functional Form of the Hedonic Price Function

In addition to specifying which variables belong in the hedonic regression, researchers must determine how the characteristics influence price. In other words, the researcher must decide on the functional form of the hedonic price function. Little theoretical guidance exists for the choice of functional form because the price schedule is determined in the marketplace by the interactions between many different buyers and sellers. If a product can be repackaged at no cost, the total price of the product would simply be the sum of the implicit prices of its component characteristics.[4] In this case, the appropriate specification for the hedonic price function would be a simple linear function:

$$P = \alpha_0 + \sum_{i=1}^{h} \beta_i H_i + \sum_{j=1}^{n} \beta_j N_j + \sum_{k=1}^{L} \beta_k L_k + \varepsilon, \tag{7.6}$$

[4]The common example here is to think of a full grocery cart as "the product" that has component characteristics that are the grocery items contained within. In this case, there is no cost to repackage the component characteristics of the grocery cart, and so the total price for the product is simply the sum of the prices for each of its component characteristics.

where P is the sales price of a house; H represents structural and property characteristics of the house, such as square footage of the living area and lot size; N represents neighborhood characteristics, such as median income and quality of schools, and could include amenities, such as ambient air quality; and L represents locational characteristics, such as proximity to the central business district, and could include proximity to environmental amenities and disamenities. In the case of the linear functional form, the implicit price for any specific characteristic, z_i, is simply the estimated coefficient for that variable ($\beta_i = \partial P(\underline{z})/\partial z_i$). With a linear hedonic price function, the incremental price of a property induced by an incremental increase in a characteristic is constant across all levels of each characteristic. This implies, for example, that the incremental value of an additional foot of lakeshore frontage would be the same for lakefront properties with 50 feet of frontage as for houses with 150 feet of frontage.

Of course, marginal prices are not likely to be constant for all characteristics, and so alternative functional forms are often used where some or all of the variables are transformed. For instance, most researchers assume that a home's price is related nonlinearly to its square footage. In particular, it is generally assumed that the total price of a house increases at a decreasing rate with square footage (see Fig. 7.1 and its associated discussion in Sect. 7.1). To capture this empirically, price may be estimated as a function of the natural log of square footage as in the Boyle et al. (1999) Maine summer home example (Eq. (7.4)), or price may be estimated as a function of square feet and a quadratic term for square feet.

A hedonic price function that allows for sales price to be affected nonlinearly by the characteristics of the property may be specified in many possible ways. Table 7.2 presents some common functional forms that introduce nonlinearity in the hedonic price function through transformation of the dependent variable, independent variables, or both. Also shown for each functional form is the specific form for computing the marginal implicit price, $\partial P(\underline{z})/\partial z_i$. As can be seen from

Table 7.2 Functional Forms for the Hedonic Price Function

Name	Equation	Implicit prices
Linear	$P = \alpha_0 + \sum \beta_i z_i$	$\frac{\partial P}{\partial z_i} = \beta_i$
Semilog	$\ln\ P = \alpha_0 + \sum \beta_i z_i$	$\frac{\partial P}{\partial z_i} = \beta_i \times P$
Double-log	$\ln\ P = \alpha_0 + \sum \beta_i \ln\ z_i$	$\frac{\partial P}{\partial z_i} = \beta_i \times P/z_i$
Quadratic	$P = \alpha_0 + \sum_{i=1}^{N} \beta_i z_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \delta_{ij} z_i z_j$	$\frac{\partial P}{\partial z_i} = \beta_i + \frac{1}{2}\sum_{j \neq i} \delta_{ij} z_j + \delta_{ii} z_i$
Quadratic box-cox*	$P^{\theta} = \alpha + \sum_{i=1}^{N} \beta_i z_i^{(\lambda)} + \frac{1}{2}\sum_{i,j=1}^{N} \delta_{ij} z_i^{(\lambda)} z_j^{(\lambda)}$	$\frac{\partial P}{\partial z_i} = \left(\beta_i z_i^{\lambda-1} + \sum_{j=1}^{N} \delta_{ij} z_i^{\lambda-1} z_j^{(\lambda)}\right) P^{1-\theta}$

*The transformation function is $P^{(\theta)} = (P^{\theta} - 1)/\theta$ for $\theta \neq 0$, and $P^{(\theta)} = \ln(P)$ for $\theta = 0$. The same transformation applies to lambda. The marginal effect is computed for $\theta \neq 0$. The marginal effect when $\theta = 0$ and $\lambda = 0$ is the same as for the double-log. The quadratic box-cox is equivalent to the linear box-cox, another common functional form when $\delta_{ij} = 0$ for all i and j. The marginal price of the linear box-cox is simply $\beta_i z_i^{\lambda-1} P^{I-\theta}$

these examples, the formulas can be complicated and can involve not only the estimated coefficients from the hedonic price function, but also the levels of the characteristics themselves.

Two possible approaches may be used for computing implicit prices when the formula involves the level of a variable. Consider the semilog functional form in which $\partial P(\underset{\sim}{z})/\partial z_i = \beta_i P$. In this case, the implicit price must be evaluated at some level of sales price. Most commonly, the mean or median housing price of the sample, over which the coefficients were estimated, is chosen. Alternatively, the implicit price may be computed for each house in the sample, and then the mean of these prices may be used. No one way is correct, although they have different interpretations. The former is the effect for the properties with the mean or median sales price, while the latter is the average effect for all sold properties. Ultimately, it is researcher judgment guided by the application at hand that will determine the appropriate level of price (or any variable) at which one evaluates the implicit price.

Product characteristics are often described by categorical variables. For instance, rather than recording the number of fireplaces in a home, a variable may be created to equal one if there are any fireplaces in the home and equal to zero if there are none. If the dependent variable is the natural log of sales price, then the interpretation of the coefficient estimate is the approximate percentage change in price when the characteristic in question is present. It is an approximation because the coefficients estimated for the dummy variables are transformations of the percentage effect (Halvorsen and Palmquist 1980; Kennedy 1981). For a coefficient estimate, b, the percentage effect, g, is given by $100 \times g = 100(e^b - 1)$. When b is small, the error in interpretation (without adjustment) is small. For instance, a coefficient estimate of 0.10 implies that the percentage change in price is actually 10.5%.

What can serve as a guide for choice in functional form? First, one should be careful not to rely on a linear hedonic (with no transformed variables) unless there are compelling logical reasons to do so. In general, nonlinear relationships between price and some size and quality attributes are expected. While researchers may use standard regression diagnostics to evaluate the fit of their chosen functional form, a classic paper by Cropper et al. (1988) takes a more rigorous approach to examining how choice of functional form may affect the researcher's ability to recover unbiased implicit prices. The authors conduct an exercise in which a simulation model of a housing market is calibrated with actual transactions data, and the resulting simulated equilibrium is used to compute the *true* marginal values for housing characteristics. These true marginal values are then compared to implicit prices computed from hedonic regressions estimated under two scenarios: one in which all housing characteristics are observed by the researcher, and one in which some variables are assumed to be unobserved by the researcher and thus are omitted from the hedonic regression. Each functional form presented in Table 7.2 was then evaluated for its ability to recover unbiased estimates of marginal values for specific characteristics. The authors found that simpler functional forms (e.g., the semilog) were better at recovering marginal values in the presence of unobserved housing characteristics. The findings of Cropper et al. (1988) have no doubt had a large impact on hedonic housing regressions over the past 25 years, as evidenced by the preponderance of semilog functional form choices made by authors, often with direct citation of Cropper et al.

More recently, Kuminoff et al. (2010) revisited the approach taken by Cropper et al. (1988) using current data and empirical methodologies, especially as relating to spatial features of housing markets. Contrary to Cropper et al., more flexible, functional forms such as the quadratic box cox are found to significantly outperform the simpler linear, semilog, and log-log specifications. The work of Kuminoff et al. has very important implications for current hedonic research, especially as researchers now consistently focus attention on controlling for unobserved spatial features of housing markets and often employ quasi-experimental methods as are described in Sect. 7.2.3.

However, note that box-cox estimation has been relatively uncommon within the hedonic housing literature. Early examples are Cheshire and Sheppard (1995), Palmquist and Israngkura (1999) and Ihlanfeldt and Taylor 2004. Given the implications of Kuminoff et al. (2010), researchers moving forward should expand their attention on flexible functional forms specifying the hedonic price function (e.g., Buck et al. 2014).

7.2.1.2 Market Definition: Space and Time

When choosing a sample of properties for a hedonic analysis, one must consider the geographic coverage of the data selected as well as the time period. First, consider the geographic coverage of the data. If a study's focus is an environmental good, then the data has to have geographic coverage sufficient to ensure variation in the environmental amenity or disamenity across properties. Depending on the amenity, variation may be in the form of the proximity of each property to the amenity (e.g., proximity to a park), or the variation may be the ambient level of the amenity (e.g., the level of air quality at the site). While variation of the first type is typically easy to ensure, ambient environmental quality can sometimes be more difficult to ensure in a sample frame. For example, in order to have sufficient variation in an environmental variable, the geographic dispersion of properties in a sample may be increased so much that properties are now drawn from different markets. In this case, estimating one hedonic price function for the entire sample is inappropriate because the hedonic price function is an equilibrium function describing a single unified market. The question that arises is how to determine what set of properties constitutes a single market. In other words, we wish to know the extent of the market.

Markets are truly separate if participants in one market do not consider houses in the other market when making purchase decisions, and sellers do not consider sales prices in the other market when negotiating sales prices. One common assumption is that each urban area represents a separate housing market. If the focus of the study is proximity to an amenity or disamenity, each urban area typically has sufficient variation for estimation.[5]

Although considering each urban area a separate market is likely a reasonable assumption, recent work has often assumed, either implicitly or explicitly, that housing markets are national (e.g., Hanna 2007; Noonan et al. 2007; Greenstone and Gallagher 2008; Grainger 2012; Sanders 2012).

In early work, Linneman (1980) considered whether housing markets are indeed national by comparing city-specific hedonic price functions for housing with a hedonic price function estimated on a national sample of housing. His hypothesis of a national market is neither fully supported nor rejected (early work is reviewed by Palmquist, 1991). The evidence for clear market segmentation across any barrier is mixed across applications, and it has often been the case that researchers do not provide evidence that supports their implicit choices.

It is difficult to test conclusively for market segmentation, but there are commonly employed methods that are used as supporting (or refuting) evidence for the researcher's priors about segmentation. The most common approach is to apply F-tests to determine if the hedonic price functions differ across segments (e.g., Taylor and Smith 2000). The problem with F-tests, however, is that results indicating that market segmentation exists may be due to mis-specification of the hedonic price function and not actual segmentation. F-tests are also likely to reject aggregation with large samples (Ohta and Griliches 1976). Once again, there are no definitive answers. However, researcher judgment along with supporting statistical tests can be used as a guide for determining the market extent in a particular study.

In addition to needing sufficient geographic variation in the data, it may be important to have appropriate time variation in the data. Often, a researcher is interested in how the valuation of an amenity changes over time. Or the researcher is interested in how prices changed in response to the creation of an amenity or disamenity, such as the siting of a landfill near a neighborhood. In some instances, the researcher may simply need observations from multiple years in order to increase sample sizes or because the research design and data dictate pooling data over many years.

When drawing a sample frame from multiple years, it is important to remember that the goal is to recover the parameters of a single equilibrium function that relates sales prices to a product's characteristics. As such, the parameters of the function should be constant over the time frame used in estimation. A single equilibrium hedonic price surface contrasts rather sharply with one's intuition about market conditions during the period 1996 to 2006 when U.S. *real* home prices had increased 86% during the decade and annual price increases of more than 10% were not uncommon in major cities around the world (Shiller 2007). However, the housing bubble burst in 2006, resulting in dramatic decreases in prices and

[5]Segmentation within an urban area has also been considered (e.g., Goodman and Thibodeau 2007).

increases in vacancy rates. Between 2006 and 2009, the U.S. housing stock was estimated to have lost nearly $4.5 trillion in value (Carson and Dastrup 2013).

Housing markets, like all markets, have cycles, and the issue is not so much one of appropriateness of using hedonic methods in cyclical markets, but rather careful understanding of how market cycles influence both estimation approaches and interpretation of results. Indeed, the dramatic consequences of the housing downturn in the late 2000s refocused the attention of hedonic researchers on how housing cycles affect best practices, especially given that modern transactions data available to researchers usually span a decade or more of sales (Boyle et al. 2012). The key question for nonmarket valuation is whether the changes in the hedonic surface are simply inflationary (or deflationary) and are easily netted out of the price function, or if housing market cycles impact the value of some housing characteristics differently than others.

If there are purely inflationary trends in a market, prices may be deflated prior to estimation by an appropriate, preferably local, price index. If an appropriate price index is not available, a series of dummy variables may be included in the hedonic regression to control for the year or quarter in which each property was sold. Either of these methods essentially control for changes in the intercept of the hedonic price function, and the assumption is that all implicit prices are changing by the same factor, and thus, adjustments to the hedonic price function intercept are all that is needed.

If underlying supply and demand conditions are changing over time, implicit prices of characteristics may change, and these changes may vary by characteristic (Cohen et al. 2014). Simple price deflators are not sufficient in this case, and the strategy taken may depend on whether there is upward or downward pressure on prices. During periods when prices are increasing, allowing for time period interactions with housing characteristics permits implicit prices to change over time and can be an appropriate strategy for capturing changing marginal implicit prices.

During cold markets, when there is downward pressure on prices, one must consider whether or not the housing market is in disequilibrium. The unique features of housing markets can lead to sticky prices after demand for housing shifts inward (Leamer 2007). In other words, excess supplies in housing markets, as indicated by higher than normal vacancy rates, can clear slowly as sellers are unwilling to adjust prices downward. Simple time-interaction variables can capture changing implicit prices during cold markets that are not severe, but if a market has substantial excess supply, as evidenced by high vacancy and foreclosure rates, additional modeling is needed to properly identify marginal prices.[6] As Coulson and Zabel (2013) discussed, both buyer heterogeneity and the composition of the stock for sale may differ in cold markets relative to normal periods. While these differences do not in and of themselves invalidate the approach of using time

[6]Note: Evidence suggests that foreclosures and vacancies can have a direct impact on neighboring property values (e.g., Campbell et al. 2011), and thus, proximity to these types of homes should be included as a disamenity in the hedonic price function to avoid omitted variables bias.

interactions to trace out changes in implicit prices over time, the problem is with interpretation of the marginal price as a representative marginal value. Zabel (2013) suggested that best practices should involve estimating the hedonic price function over a full housing cycle and then averaging the estimated implicit prices across the time period to compute an average implicit price for use in policy analysis.

How might one test for stability in the hedonic function? While F-tests may be applied to test for the stability of the hedonic function and the marginal prices of individual characteristics over time, the same weaknesses in the tests applies as stated previously for tests of market segmentation. Ohta and Griliches (1976) suggested a less stringent rule in which the standard errors for the constrained and unconstrained regressions are compared, and aggregation over time is rejected if errors increase by more than 10% (Palmquist 2006). As noted, substantial excess supply can be indicated by unusually high vacancy rates, and similarly, excess demand is measured by vacancies below the natural rate. The researcher should understand the market dynamics during the sample frame of the data and be aware that relying on data from housing markets that are making large adjustments is not likely to yield representative results for the implicit price of an environmental amenity (Zabel 2013).

7.2.2 Data Sources, Quality, and Measurement

A perfect dataset—ready for analysis and designed to answer a specific question that a researcher is interested in exploring—does not exist. The researcher will need to obtain data from multiple sources, each with their own challenges, and integrate them into a single database suitable for the analytical tasks at hand. This section explores a few commonly used types of data and highlights key challenges presented by the data for estimation of hedonic price functions.

7.2.2.1 House Value and Characteristics

Housing sales prices and characteristics are available from many sources. In many areas, the local tax assessor's office maintains electronic databases of their tax rolls that include a unique identifier for each parcel, its most recent sales price, lot size, a description of improvements (buildings) on the parcel, and other useful information, such as the school district to which the parcel is assigned (see Phaneuf et al., 2013, and Zabel and Guignet, 2012, for examples of microdata obtained from tax assessors).[7] These records are maintained at the county, city, or some other level of

[7]It can be the case that one residential property is formed by more than one parcel, and this is often true with commercial properties. The researcher must be careful that the final data used for analysis properly represents all the land (and buildings) contained in a single sale and that there is only one observation in the data for each unique sale.

municipal jurisdiction. Private vendors also collect local real-estate transactions data for re-sale in electronic formats (see Cohen et al., 2014, Kuminoff and Pope, 2013, Bajari et al., 2012, and Taylor et al., 2016, for references to sample vendors). While these data are often quite rich, they must be purchased. This contrasts with tax assessor records that are typically available free of charge or for nominal fees. For example, the King County tax assessor's office, which encompasses Seattle, Washington, provides rich data on sales prices, property characteristics, and locational characteristics via free download from its website.[8]

Data purchased from private vendors can have advantages over that obtained from tax assessors. First, privately provided sales data often includes all sales of each property over the time period for which data are provided. This can allow repeat sales models to be estimated, which is not possible when public data records only the most recent sale price for a property (e.g., Cohen et al. 2014). Second, assessor data may not be updated for property characteristics, whereas private data is often based on the National Association of Realtors multiple listing service data, which should be current. A shortcoming of private vendors is that data is often collected for metropolitan areas only. Tax assessor data are available for every location, including rural areas, although the quality and format in which the data are available may vary greatly.

Whether purchased from a private vendor or obtained directly from a municipality, records of actual transactions prices coupled with property characteristics are the norm in hedonic research. Once obtained, the researcher should carefully assess the data's quality.

First, sales prices may be recorded in error, resulting in unusually high or low sales prices. Even if sales prices are recorded without error, not all sales are arms-length transactions. In some cases, properties may be transferred to family members or from one business entity to another at prices much less than market value. These observations would not be appropriate for inclusion in the data set because the hedonic price function is an equilibrium function arising from competitive market bidding for each property. Unfortunately, recognizing transactions that were not arms-length is not always easy. An intuitive approach is to omit any observations on prices that are implausibly low or high. However, the researcher must carefully define what is too low or too high and should have a good understanding of the market.

Another consideration for sales data is the potential for sample selection bias. Here, the concern is that properties with specific unobservable characteristics may be less likely to be placed on the market for sale or less likely to sell once placed on the market. Sample selection bias may be particularly important in studies focusing on housing that is proximate to amenities or disamenities. For example, unobserved characteristics that influence the likelihood a home is placed on the market or is sold could be correlated with proximity of a house to a hazardous waste site or other disamenity. In this example, the ordinary least squares (OLS) estimate of the

[8]See http://info.kingcounty.gov/assessor/DataDownload/default.aspx; last accessed Nov. 5, 2013.

implicit price for distance from a hazardous waste site will be a biased estimate of the true market price, and the direction and magnitude of the bias will generally be unknown. It is not known if an explicit consideration of this issue has ever been undertaken. In a related paper, however, Huang and Palmquist (2001) developed a model in which they explicitly consider the simultaneity between the duration a property is on the market and its sales price. They argued that housing located near a disamenity may suffer two impacts: reduced sales prices and reduced probability of sales. They found that location near a noisy highway does not adversely affect duration on the market, but does adversely affect sales prices. However, they did not test for selection effects among the mix of properties offered for sale nor for bias in implicit price estimates, and this is an area where further research is warranted.

Finally, transactions data may be sparse and thus may need to be aggregated over a large number of years in order to have enough observations to reliably estimate implicit prices, especially as related to specific neighborhood or locational characteristics. However, aggregating over significant periods of time can create its own challenges, as was discussed in Sect. 7.2.1.2.

The use of survey data can overcome some of the limitations that transactions data may present. For instance, when research designs call for broad geographic coverage spanning multiple states and transactions data are prohibitively expensive to obtain, the researcher may use survey microdata collected by the U.S. Census Bureau.[9] Two types of microdata are available. The first is collected as part of the decennial census. In census years 2000 and earlier, the U.S. Census Bureau conducted a 1-in-6 sample of every household in the U.S. (approximately 18 million housing units in 2000), and these households were asked to answer what is referred to as the long form. The long-form survey asked a series of questions regarding housing that can be useful for hedonic researchers, including home value and basic housing characteristics.

Census survey microdata is publicly available. In other words, the exact survey responses for every individual who fills out a census form is made available to the public. However, the location of the household is confidential information, so the Census Bureau only releases survey results that identify the household's location by public use microdata area. Each public use microdata area is a statistical area that comprises at least 100,000 people.[10] Unfortunately, many, if not most, environmental amenities considered in hedonic housing studies vary at smaller scales than public use microdata areas, rendering the data incapable of estimating implicit prices for amenities and disamenities that vary substantially within a public use microdata area.[11] To address this limitation, some researchers are taking advantage of Research Data Centers through which the Census Bureau will grant researchers special access to the survey microdata that includes information on the city block in

[9]Microdata here refers to the individual survey responses.

[10]See www.census.gov/geo/reference/puma.html for more information.

[11]Bajari and Kahn (2005) used census public use microdata area in work analyzing racial segregation within cities.

which each home is located (see Gamper-Rabindran and Timmins, 2013, Liu et al., 2013, and Davis 2011).[12]

Finally, while individual survey responses linked to a home's location are not publicly available, the Census Bureau does publicly report aggregated housing statistics, such as median owner-reported housing value for census-defined neighborhoods. The smallest geographic area for which the Census Bureau releases median house values and housing characteristics is neighborhoods defined by "block groups," which can contain as few as 600 people but can contain up to 3,000 people. Block groups are meant to represent neighborhoods of similar people (see Noonan et al., 2007, for an example hedonic study using block groups).[13]

Researchers also use data aggregated by census tracts (e.g., Hanna 2007; Greenstone and Gallagher 2008). Census tracts are aggregations of block groups and usually contain between 1,600 and 8,000 people, although the Census Bureau indicates that 4,000 people is an optimum size for a tract. Census tract boundaries typically follow identifiable features (e.g., major roads) and do not cross county boundaries.

Census tract and block group summary data are attractive to researchers because of their ease of availability and universal coverage. However, given the aggregated nature of the data, the researcher has to pay close attention to the relationship between the spatial variation of the amenity of interest relative to the geographic scale of census block groups or tracts. When environmental amenities are expected to have external impacts that vary at a scale smaller than tracts or block groups, it can be difficult or impossible to identify their external effects. An excellent discussion of this point and how the researcher can address the problem within the context of publicly available census data is presented in Gamper-Rabindran and Timmins (2013).

A potential shortcoming of census data is that the Census Bureau discontinued the long form in 2000 and replaced it with the American Community Survey, which is a continuous survey that samples approximately 3 million addresses each year. A single sample does not fully cover the U.S. and is not likely to have enough observations in small geographic areas for valuing localized amenities. However, the Census Bureau samples in a way that the aggregate of samples over each contiguous five years has complete coverage of the U.S. and mimics the previous census long-form data in content.[14] The quality and appropriateness of this data for hedonic nonmarket valuation studies—especially in light of housing cycle impacts on implicit price estimates—is an area of research that has yet to be explored.

---

[12]See www.census.gov/ces/rdcresearch/ for more information.

[13]The smallest geographic area the Census Bureau defines is a census block, which is an area defined by streets or natural boundaries (e.g., imagine a city block in an urban area). Houses are identified by the block in which they are located only through special access to Census Bureau Research Data Centers. Census blocks are aggregated to block groups, usually containing approximately 40 blocks.

[14]See www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf for details on the American Community Survey. Last accessed Nov. 5, 2013.

An additional shortcoming of using census survey microdata is the potential for measurement error in the dependent variable because surveys ask individuals dwelling in a unit to estimate the value of the property. This is only a problem if the measurement error is systematically related to observed or unobserved characteristics of parcels. There are few studies that have systematically compared homeowner-assessed values to transactions data, but when they have, the results suggest little systematic bias in reported values.[15]

7.2.2.2 Neighborhood and Locational Characteristics

Most environmental amenities that researchers focus on in hedonic housing models are spatial in nature. Either the amenity directly varies over space (e.g., ambient air quality or density of tree cover) or the amenity is fixed in place and the researcher measures "exposure" of a home to the amenity through its spatial proximity (e.g., proximity to open space). Usually, measures of a home's exposure to a local amenity or disamenity have to be developed by the researcher directly. To do this, the first step is to develop a geographic information system database that contains accurate geospatial references for each property and each amenity of interest. Geospatial references are usually either boundaries of a feature, such as the boundary of a state park, or the center point of a feature, such as the center point of the central business district. Typically, the researcher will create a geographic information system database for properties in an area of interest and then merge this data with existing data obtained from municipalities, environmental agencies, or other entities that have data on environmental or locational characteristics of interest. By merging these two geographic information system data types together, the researcher creates a complete geospatial dataset with which to define (measure) a set of neighborhood and location characteristics for inclusion in estimation of the hedonic price function. Each of these steps is described more fully in turn, along with key considerations that can impact how an analysis is conducted.

If transaction microdata are collected from tax assessors' offices, one can usually also obtain a digitized tax map that contains the boundaries for every property in geographic information system format. These tax maps will identify each property by its unique parcel identification number, and that number can be linked back to the microdata recorded by the tax assessor, such as the most recent sales price of the property. If a digitized tax map is not available, street addresses can be used to

[15]Kiel and Zabel (1999) provided a comparison of owner assessments with actual sales prices for three metropolitan areas over approximately an 11-year period. Their results indicate that although owners generally overestimate the value of their homes by approximately 5%, the difference in owner valuation and actual sales price is not related to any housing or owner characteristic other than the tenure in the home (those living longer periods of time in their house at the time of the survey provided more accurate values). Although not a direct comparison, Gamper-Rabindran and Timmins (2013) compared overall results of an analysis conducted with national census data to parcel-level transactions data and find consistency across the two analyses.

assign latitude and longitude coordinates to the property using readily available software, such as ArcGIS. Address matching is not as precise as using a tax map because the location of a property is only approximate along the street segment in which the property is located.

Given a geographic information system database for properties, the spatial relationship between any one property and surrounding features of interest may be developed. Geospatial data for political boundaries (states, counties, municipalities) and transportation networks are available nationally from the U.S. Census Bureau (www.census.gov/geo/maps-data/), and they are also embedded within geographic information system software packages. Federal, state, and local agencies maintain rich datasets that are usually downloadable for free or for a very modest fee. Most university libraries also subscribe to geospatial data vendors that gather data from multiple sources and provide it in user-friendly formats. Luckily, rich geospatial data is typically available from urban county or city governments, planning agencies, and state and federal environmental management agencies. With a bit of searching, it is possible to build a rich geographic information system database of locational features for a study area.

An important source of neighborhood characteristics is the U.S. Census Bureau. Through the decennial census or American Community Survey, the Census Bureau publicly releases summary statistics of neighborhood characteristics, such as the percentage of owner-occupied homes and the racial composition and age distribution within a neighborhood. These data are released publicly at the census block group or tract level and are quite useful for measuring neighborhood characteristics.

The researcher can use block group or tract data in two ways. First, they can directly include summary statistics that capture neighborhood features of interest, such as median income or the percentage of owner-occupied homes in a neighborhood. Second, if the researcher is using sales price microdata, block group or tract-fixed effects can be included in the regression to capture all time-invariant characteristics associated with the home's neighborhood.[16]

Creating measures of environmental amenities is not necessarily a straightforward task. Ideally, one would have each characteristic measured in a manner consistent with the perceptions or understanding of the characteristic by the market participants. For instance, one can imagine many ways to construct a measure of proximity of a home to a nearby amenity. Linear distance of the center of a residential lot to a feature of interest is a common measure.[17] However, as Matthews and Turnbull (2007) indicated, measuring distance by the street network may be more important for highly localized amenities. In addition, proximity may not capture all the relevant features of "perceived exposure." For example, most studies

[16]When not using census tracts or block groups, researchers may use subdivision names, school districts, or local jurisdictions as fixed effects.

[17]Another common measure of proximity to a specific site or feature of interest is the use of buffer zones, such as a series of dummy variables that indicate if a property is within a certain distance band of a feature of interest (e.g., Taylor et al. 2016).

exploring the impact of hazardous waste sites on property values focus on the distance between a home and the nearest site.

Given agglomeration of land-use types, it is often the case that homes are proximate to multiple sites with varying intensity of environmental contamination (see also Leggett and Bockstael, 2000). The density of contaminated sites coupled with the intensity of contamination at each site is rarely addressed in studies. Furthermore, there may be several ways to measure the level of contamination present at a site, and it is unclear what measures best reflect market perceptions and, thus, are relevant to include in the hedonic price function.

If there is a high degree of correlation among measures relating to the same characteristic, including them all may lead to imprecisely estimated coefficients. On the other hand, dropping a variable to avoid multicollinearity could very well introduce bias. These are classical econometric issues. Clearly, balanced judgment on the part of the researcher, as well as sensitivity analyses on the choices made, are all part of the well-conducted hedonic study.

Measurement error in the constructed variables describing environmental amenities or disamenities is not limited to construction of spatial variables. For example, ambient environmental conditions such as air quality at the site may also be difficult to quantify in a manner that accurately reflects buyers' and sellers' perceptions of the characteristic. In Boyle et al. (1999), the Maine Department of Environmental Protection provided the measure of lake water clarity used in the hedonic regression analysis, and the clarity was measured monthly at a specific site on each lake during the summer months. Prospective purchasers may not have visited the lake during those months (and water clarity varies throughout the year on Maine lakes). As such, the objective or scientific measures of clarity collected by the Department of Environmental Protection may not have coincided with water clarity as perceived by prospective property purchasers. In a later analysis of the same market, Poor et al. (2001) found that an objective measure of water clarity at the time of sale was either preferred, or equally preferred, to the homeowner's subjective assessment for explaining variation in sale prices.

Although scientific measures or the proxy variables used by researchers are hoped to be reasonably well correlated with the *true* factors that are capitalized into a property's value, the use of proxy variables can nonetheless introduce the potential for errors-in-variables problems, leading to biased estimates of all coefficients. If only one variable is measured with error, improving the measurement of that variable will reduce the bias in all coefficients. However, when more than one variable is measured with error, reducing measurement error in one variable will not necessarily reduce the bias in the other regression coefficients (Garber and Klepper 1980).

Again, no definitive answer can be offered with respect to building the data needed for the hedonic model. Common sense, research and knowledge about the study market, and an understanding of the consequences of the choices (assumptions) made are key to conducting a thoughtful and careful hedonic price study. If multiple measures are available, the researcher should favor those that are consistent with economic intuition about what is likely to best represent the externality of interest as perceived by market participants, and report the sensitivity of the results to variable construction choices.

7.2.3 Endogenous Regressors

Endogeneity in hedonic models has long been a prominent research theme. Endogeneity can arise from simultaneous determination of sales price and a regressor or from omitted variables. In either case, the result is that coefficient estimates will be biased. Endogenous regressors are a classical econometric problem, and while the solutions follow classical techniques, there have been specific approaches focused on in the hedonic literature that are discussed here. The econometric approaches used to identify unbiased coefficient estimates are organized by the source of endongeneity: simultaneity or omitted variables.[18]

7.2.3.1 Simultaneity

First, consider endogeneity that arises from joint determination of sales price and a regressor. An example of this relates to land-use spillovers, which is particularly important for studies that seek to understand the value of proximity to open space for residential homeowners. Following Irwin and Bockstael (2001), consider the case in which all land is privately held and open space is potentially developable, as is typically the case at the urban fringe. For simplicity, imagine two properties adjacent to each other: parcels *i* and *j*. The value of parcel *i* in time *t* is a function of the amount of open space surrounding it—i.e., the open space on parcel *j*—and vice versa. However, the amount of open space around parcel *i* is an economic function of the value of its neighbor, parcel *j*, in development, and vice versa.

Addressing endogenous regressors that arise from joint determination relies on instrumental variables approaches (see Irwin and Bockstael, 2001, for a thorough discussion and empirical example). In other words, the researcher must find variables that are correlated with the endogenous variable (amount of open space surrounding a property in the above example) but uncorrelated with the error term. In considering open space around a particular property, Irwin and Bockstael used the characteristics of the land of neighboring properties (e.g., soil quality and land slope) as an instrument for the open space surrounding a particular property because these features serve as a proxy for the cost of developing the land.

[18]In this chapter and in the hedonic literature, two types of identification are discussed: identification of unbiased coefficient estimates and identification of the demand function for a characteristic of interest. Research that only estimates a first-stage hedonic price function would only discuss the former type of identification, while research that includes a second-stage analysis as presented in Sect. 7.4 is likely to include a discussion of both types of identification.

It is important to note that endogeneity due to the simultaneous determination of price and a characteristic of housing does not arise in a first-stage hedonic regression because housing prices are modeled as nonlinear functions of their characteristics. For instance, consider a commonly included characteristic—square footage of a home. Prices are usually modeled as varying nonlinearly with increased square footage. This makes intuitive sense because one would not expect an extra square foot of space to have the same marginal value for a 4,000-square-foot home as it does for an 800-square-foot home. However, square footage is not endogenous in a first-stage regression because of the fact that households that choose larger homes pay lower (or higher) marginal prices, ceteris paribus. The underlying parameters of the demand or supply of square footage is not being estimated in the first-stage regression, and as such, it is not endogenous. Of course, endogeneity of square footage could arise for other reasons. For example, square footage might be correlated with an important omitted variable, and if this is the case, the endogeneity would need to be addressed as described in the next section.

7.2.3.2 Omitted Variables: Spatial Econometrics and Quasi-Experiments

The second source of endogeneity being considered arises from omitted variables. As noted previously, housing location plays a critical role in most environmental valuation applications. As such, an important concern in hedonic housing research is that parameter estimates for spatially varying environmental amenities could be correlated with unobserved spatially varying characteristics.

Concerns over spatially varying omitted variables generally fall into two distinct categories. In the first case, one may believe that omitted spatial variables are independent of the regressors and only introduce spatial error correlation. In this case, spatial econometric methods may be introduced. Following Anselin (1988), a spatial error model may be written as follows:

$$\begin{aligned} P &= ZB + \varepsilon, \\ \varepsilon &= \lambda W\varepsilon + \mu, \\ \mu &\sim N\left(0, \sigma^2 I\right), \end{aligned} \tag{7.7}$$

where P is sales price, Z is a $N \times K$ matrix of property characteristics, B is a $K \times 1$ vector of coefficients, W is an $N \times N$ spatial weights matrix, ε is a $N \times 1$ spatial autoregressive error, μ is a $N \times 1$ random error term with variance σ^2, and λ is the coefficient on the spatially lagged error term ε. In the above, B and λ are estimated, and W is chosen by the researcher. The equations in (7.7) may be rewritten as:

$$P = ZB + (I - \lambda W)^{-1}\mu. \tag{7.8}$$

Specification of the spatial weights matrix, W, continues to be one of the more controversial aspects of spatial econometrics because results are sensitive to the choices made. The spatial weights matrix defines the sense in which properties are believed to be neighbors and determines the importance of any one observation to another. It is similar to a lag operator in the analysis of time series data but is multidimensional. In housing hedonics, distance-decay matrices are the most common specification of the weights matrix, wherein the importance of each property on the current property decays as distance increases. Alternative structures such as lattice matrices are also used. Lattices may be specified in several ways, such as allowing elements of the spatial weights matrix, w_{ij}, to equal one if a property shares a border with the observation of interest and equal to zero otherwise. Or, weights may be equal to one if properties are next to the property of interest or are contained within some predefined radius of the property of interest.

The second type of concern with regard to omitted variables is when they are thought to be correlated with other regressors, and classical endogeneity problems arise. In this case, parameter estimates will be both biased and inefficient. Two approaches are commonly employed to alleviate omitted variable bias. In the first, a spatial lag term (or spatial autoregressive term) may be added. Specifically, the model in Eq. (7.8) may be extended to allow for a spatial autoregressive process as follows:

$$P = \rho W_1 P + ZB + (I - \lambda W_2)^{-1}\mu, \tag{7.9}$$

where we now specify two weights matrices, W_1 and W_2, that are allowed to differ, and both ρ and λ are coefficients on the spatially lagged variables. If ρ and λ are estimated to be equal to zero, Eq. (7.9) reduces to a simple linear regression model. In a spatial autoregressive model, neighboring values are modeled as partially determining a home's value. This approach can be used to ameliorate the effect of omitted variables by using lagged prices as a proxy for unobserved neighborhood characteristics that contribute to the utility of owning a particular home in that neighborhood. The strength of this approach will rely on proper specification of W_1 and use of appropriate instruments for lagged prices because they too are endogenous.[19] Brady and Irwin (2011) provided an excellent, concise review of current best practices in applying spatial econometric models to common questions in environmental economics, while LeSage and Pace (2009) and Anselin (1988) provided technical introductions to spatial econometrics. Routines to implement spatial regression models are available in free statistical software, such as GeoDa or *R*, or proprietary software such as SpaceStat.

While the spatial autoregressive model can be a tool to recover unbiased parameter estimates for key variables of interest, there has been some confusion as

[19]While spatial autoregressive models can be employed to alleviate omitted variables, it should be noted that they may be used to directly model simultaneity as described in Sect. 7.2.1.1 (Irwin and Bockstael 2001).

to the economic intuition underlying the use of a spatial autoregressive model. Some have suggested that the intuition is that the spatial autoregressive model captures the price *determination* process rather than invoking a straightforward omitted variables argument. It is often suggested that because Realtors look to nearby property sales prices for "comps" (comparable homes) to help set the offer price for a home, these neighboring sales prices partially determine (influence) the sales price of a particular home. However, this is a mischaracterization of the price determination process. It is true that Realtors and homeowners use sales prices of nearby similar homes as an information tool to discover the nature of the equilibrium price surface and use this information to develop an initial offer price. If one believes the assumptions of the hedonic model, the ultimate sales price is determined by the interactions of many buyers with many sellers of competing homes in a competitive market. If a homeowner or Realtor misinterprets the information contained in nearby sales transactions and sets an offer price too far above the equilibrium for a home with its specific set of characteristics, the home will go unsold unless the owners are willing to reduce the price.

Note that the above argument does not indicate that a spatial autoregressive model should not be used. In fact, it can be an important means by which researchers can ameliorate potential biases in parameter estimates of interest that arise from omitted variables. The importance of the above point is highlighted when considering whether a spatial multiplier should be used to compute the benefits of an amenity improvement.

Briefly, a spatial multiplier approach indicates two benefits should be added together. The first is the direct benefit that an amenity improvement has on a property as measured by the marginal implicit price discussed in Sect. 7.1. To this direct effect, an indirect effect is added that is the impact of the amenity change on the price of the home in question through the channel of the autoregressive term in Eq. (7.9), $\rho W_1 P$. This latter effect occurs because the amenity change will also increase the price of neighboring homes (see Kim et al., 2003). However, as Small and Steimetz (2012) discussed, this indirect effect is only appropriate to apply if the *utility* of the homeowner is directly affected by the price of neighboring homes. In other words, if homeowners suffer from a type of money illusion and utility is impacted by the sales price of nearby homes irrespective of the real characteristics that underlie the prices, then a spatial multiplier is appropriate to add to benefits estimates.[20] This sort of pecuniary motivation seems unlikely.

---

[20]Another way to see this issue is to consider the reason why, or why not, lagged sales prices should be in the hedonic price function to begin with. This point is best explained with a simple example. Assume Home A is for sale. Home B is located near A, and was sold recently. Home B's characteristics may affect the utility of prospective buyers of Home A (e.g., attractive landscaping and home style). Inclusion of a lagged price after controlling for *all* real characteristics of nearby properties that affect the homeowner's enjoyment of his or her own home indicates that he or she receives utility just from the fact that a nearby home sold for a certain price. A pecuniary motivation for paying a higher price for Home A in this case indicates that people suffer from a type of money illusion or have speculative motives, neither of which is appropriate to include when calculating benefits of amenity improvements.

A second approach to alleviate omitted variables bias that has received growing attention in the hedonic literature is the use of quasi-experimental research designs. A quasi-experimental design broadly refers to a study design that exploits a transparent, exogenous source of variation in the explanatory variable(s) of policy interest that divides the properties into "treated" and "untreated" types. An example of treated properties could be homes that are located next to a hazardous waste site that is slated for cleanup due to a government program.

In the parlance of program evaluation literature, the evaluation problem is that researchers observe households in one state—treated or not—but never both (Moffit, 1991, and Meyer, 1995, provided concise introductions to program evaluation). To identify the effect of a treatment on the treated observations (e.g., the effect of hazardous waste removal on the price of nearby homes), valid counterfactuals for the treated observations need to be constructed. Researchers need to determine what the outcome of the treated group would have been had they not been treated. In classic experiments, researchers randomly assign individuals to treatment and thus can expect the outcomes of the control group—had they actually been treated—to be equivalent to the outcomes of the treated group up to sampling error. Through randomization, the control group is expected to provide a valid counterfactual for the treated group—they provide an estimate of the outcome of the treated group had they *not* been treated.

In hedonic modeling, the researcher is generally not able to randomize houses to a treatment. Instead, the researcher must rely on a naturally occurring process, such as a natural event or a policy decision that assigns homes to treatment in an exogenous manner (i.e., in a way that is unrelated to the error term in the hedonic regression). For this reason, quasi-experiments are also referred to as natural experiments. For example, Petrie and Taylor (2007) exploited a natural experiment to estimate the value of access to irrigation for agricultural land. They exploited a policy decision made by state regulators who issued a (surprise) halt to the issuance of irrigation permits in a particular watershed. Irrigation permits had previously been essentially free and available to all landowners that applied. After the policy change, the only way to acquire a new permit was to purchase land that had already been granted a permit (permits were not transferable otherwise). The policy change thus turned a free asset into a scarce asset whose value should be reflected in sales prices of agricultural land with permits. It was reasonably argued that the policy applied to agricultural parcels in a way that was unrelated to factors that determine land prices, and properties just outside the policy watershed boundary served as a control group. The authors could then exploit both time (pre- and postpolicy change) and space (inside and outside the watershed) to identify the value of access to irrigation by estimating the implicit price of permits in a hedonic regression model.

When exploiting a natural experiment to identify the value of an amenity, the process by which individuals (or households in this case) are selected into treatment defines the tools needed to identify a treatment effect. If selection is solely determined by observable characteristics, then matching estimators may be employed (Todd, 2008, provided an excellent overview of matching estimators). While this approach has received some attention in the hedonic housing literature (see, for example, Liu and Lynch, 2011), it has been more common to employ methods that address selection based on unobservables.

As discussed previously, a primary concern has been unobserved spatial characteristics that might correlate with amenities of interest. If there is selection into treatment based on unobservable differences between the treatment and control groups, estimated treatment effects will be biased. The two main approaches for alleviating this problem are to employ difference-in-difference designs or regression-discontinuity designs. Parmeter and Pope (2013) provide an excellent introduction to quasi-experimental research designs as applied in hedonic housing research.

To describe the most common difference-in-difference model in a hedonic housing context, consider a situation in which the researcher has available repeated cross-sectional data, and there is a policy that randomly assigns some homes to the treatment status. Following Parmeter and Pope (2013), a difference-in-difference model in this case can be written as

$$\ln(P_{ijt}) = \alpha_0 + \alpha_1 d_t + \alpha_2 d_j + \beta_0 d_{jt} + \sum_{k=1}^{K} \beta_k z_{ki} + \varepsilon_{ijt}, \tag{7.10}$$

where i denotes the observation, j denotes the group ($d_j = 1$ if treated, $d_j = 0$ if not treated), t denotes the time period ($d_t = 1$ after the policy change, $d_t = 0$ before the policy change). Thus, $d_{jt} = 1$ if the property is treated and the sale is after the policy change; otherwise, $d_{jt} = 0$. All other characteristics that determine sales prices are included in the vector z. In Eq. (7.10), pure time effects are captured by α_1, and pure group effects are captured by α_2. The effect of the treatment on the value of the treated houses (those affected by the policy) is given by β_0. To see this, consider the example above and say that a researcher has a sample of all home sales in a market within one-half mile of a contaminated property. In this case, $d_j = 1$ if a home is within one-half mile of an environmentally contaminated property that receives cleanup, and β_0 captures the increase in house price due to cleanup of a nearby contaminated site—the treatment effect of interest.

Note: For reasons discussed in Sect. 7.3, this treatment effect is referred to as the capitalization rate (Kuminoff and Pope 2014). The parameter α_1 captures any time-trend differences pre- and post-policy across the entire sample, while α_2 captures any time-invariant differences between housing surrounding sites that are cleaned up versus those surrounding sites that are not cleaned up.

A related approach to a difference-in-difference model is the regression-discontinuity design that can be employed when the researcher has a single cross-section available. In a regression-discontinuity design, assignment to treatment is determined fully or partly by the value of a covariate relative to a threshold. In the simplest formulation, there is a deterministic function of a single covariate known as the forcing variable, D, where $D_{1i} = 1$ if and only if $z_i \geq c$. Here, all z_i with a value of at least c are assigned $D_1 = 1$. This sharp design gives rise to a model identical to that in (7.10), except that rather than time, t, there is a different boundary. For example, the model may be written as

$$\ln\left(P_{ijs}\right) = \alpha_0 + \alpha_1 d_s + \alpha_2 d_j + \beta_0 d_{js} + \sum_{k=1}^{K} \beta_k z_{ki} + \varepsilon_{ijs}, \tag{7.11}$$

where all is as defined in Eq. (7.10) except that s denotes a spatial boundary (d_s = 1 if a home lies within the boundary, and d_s = 0 otherwise). In this instance, one exploits discrete, exogenous variation over space. In a regression-discontinuity design, one limits the sample to those influenced by the forcing variable (close to the boundary) to alleviate omitted variable bias in the s and j dimension. For instance, in a non-environmental context, homes on either side of a school attendance boundary could be used to estimate the value homeowners place on school quality (see Black, 1999).

In hedonic applications, the above models exploit discontinuities in time and/or space to identify treatment effects. Several types of quasi/natural experiments have been exploited in this context. One type takes advantage of sharp changes in information available to homebuyers. Changes in seller disclosure laws were used by Pope (2008) to identify the value of reducing airport noise. Davis (2004) used a sudden, unexplained increase in cancer cases in a community (and the associated news coverage) to value reductions in cancer risks. These studies highlight how changes in governmental information about local conditions can provide opportunities for the researcher to examine how locational disamenities are capitalized into housing prices.

A number of studies have used natural events such as hurricanes to identify the value of changes in flood risk perceptions (e.g., Hallstrom and Smith 2005; Bin and Landry 2013) or a nuclear transit accident to identify the value of avoiding risks associated with being close to a transportation route (Gawande et al. 2013). Perhaps the most common type of application in hedonics has been to take advantage of policy changes that induce quasi-random assignment to treatment. Here, treatment status—such as being located near a hazardous waste site that is remediated or being located in a watershed where new irrigation permits are no longer available—is determined by known factors that serve as valid instruments. A few examples include Chay and Greenstone (2005), who used changes in the Clean Air Act to value air quality; Petrie and Taylor (2007), who used a change in a state's regulatory permitting process to value access water for agricultural production; and Greenstone and Gallagher (2008) and Gamper-Rabindran and Timmins (2013), who valued hazardous waste site cleanup by exploiting particular features of the rules used by the U.S. Environmental Protection Agency for determining what sites would receive funding for cleanup.

Although natural experiments are a useful strategy for addressing endogeneity concerns, they too have underlying assumptions and data vulnerabilities that do not make them a panacea for identifying unbiased treatment effects. First—and key among the conditions that must be met to identify treatment effects in all applications—is that selection into the treatment must not depend on unobservables. This condition can be hard to ensure in housing applications because it is often the case that broad samples (across time or space) are needed to gather enough housing sales observations for analysis. Other underlying assumptions that are critical to understand and test their validity will vary by approach and model specification. For instance, Eq. (7.10) assumes that implicit prices for all other characteristics, β_k, are equal across groups. If the β_k is not equal across j, then one might worry about unobservable characteristics that vary across j as well. In addition, one assumes the samples are selected from the same population in the same way over time and across groups. This might be violated if, say, the change in policy results in a different mix of housing being placed for sale postpolicy.

Another very important consideration for nonmarket valuation is that the situations and data that allow one to implement a quasi-experimental approach may result in estimated treatment effects that only reflect capitalization rates but do not reflect underlying willingness to pay for the change in the amenity (Kuminoff and Pope 2014). This point is discussed more fully in the next section that develops the set of welfare theoretic measures of net benefits from an amenity change that can be computed by estimating a hedonic price function.

7.3 Welfare Measurement with the Hedonic Price Function

Our model indicates that the implicit price of amenity $i(\rho_{zi} = \partial P(\underline{z})/\partial z_i)$ is equal to the consumer's marginal WTP for the amenity $(\theta z_i = \partial \theta(\underline{z}, y, U)/\partial z_i)$. Implicit prices are the most commonly reported result from hedonic studies. If one is interested in whether or not a current stock of an amenity is capitalized into the market for housing, estimates of the marginal implicit prices (sign, magnitude, and statistical significance) are the appropriate measures to report. In the context of expenditures on other aspects of housing, these relative prices can provide interesting insights about the importance of various housing amenities. Of course, it should be recognized that if an amenity change affects consumers in other ways (say, increases the recreation benefits to consumers who do not own nearby property but visit the park after it is improved), these benefits would not be captured using the measures discussed in this chapter.

In instances where we want to know the value consumers might place on a *change* in an environmental amenity, the relationship between implicit prices and appropriate measures of WTP for the change depends on the situation. To examine this issue, this section considers owners of properties and renters of properties separately. Even though the owner and renter can be the same person (implicitly renting from oneself), this discussion considers these two sides of the market separately. Two types of changes are considered: a change in a localized amenity and a change in a nonlocalized amenity. Examples of localized externalities that have been studied in the hedonic context are highway noise (Palmquist 1992a), a hazardous waste site (Kohlhase 1991), and an incinerator (Kiel and McClain 1995). In these cases, a change in the amenity affects a relatively small number of properties, so the equilibrium hedonic price function for the entire market is unaffected. When changes in nonlocalized amenities, such as the air quality of a city occur (Zabel and Kiel 2000; Sieg et al. 2004), the amenity change affects a large enough number of houses to reflect a marketwide shift in supply, and thus, one would expect a change in the market-clearing equilibrium hedonic price function.

For a marginal change in an amenity is localized, first consider the effects on renters of the property if no transaction costs are associated with moving. When a decrease of an amenity occurs, the renters are no longer at their optimal bundle of housing given that they face the same hedonic price schedule as before the change at their home. If renters can move without cost to a house with the original bundle of characteristics, there will be no change in welfare for the renters. However, the owners of the property realize a capital loss on the property because of the decrease in the amenity level associated with the property. The owner would be willing to pay an amount of money up to the value loss of the property value to avoid the amenity change. If, in addition to being a localized change, the amenity change is marginal (i.e., a one-unit change), then willingness to pay is simply the implicit price, $\partial P(\underline{z})/\partial z_i$, for each property owner. The total willingness to pay is the sum of the implicit prices across property owners who receive a change in the amenity.

If the amenity change is localized and nonmarginal in the magnitude of change, owners would be willing to pay an amount equal to $P^1 - P^0$, where P represents the sales price of the property with the initial level of the amenity (P^0) and the new level of the amenity (P^1). The total WTP is the sum of the price changes for houses that receive the amenity change. If the analysis is done prior to an environmental change taking place, as is typical for policy analyses, the price change is forecast using the estimated hedonic price function.[21] Thus, for a change in an amenity, say characteristic z_1, the total change in welfare or net benefit (NB) is computed in the following way:

$$\mathrm{NB} = \sum_{k=1}^{N} P_k\left(z_{1k}^1, z_{2k}^0, \ldots, z_{nk}^0\right) - P_k\left(z_{1k}^0, z_{2k}^0, \ldots, z_{nk}^0\right), \tag{7.12}$$

where $P_k(z_{1k}, z_{2k}, \ldots, z_{nk})$ is the hedonic price function evaluated with the kth property's characteristics in either the initial or new state.

For example, using the estimated hedonic price function in Eq. (7.1), a property with characteristics equal to the sample means (see discussion following Eq. (7.1)

[21] In the case of a localized amenity, researchers are able to forecast the change in price prior to the amenity change actually occurring because the hedonic equilibrium does not change. One still must be careful that the magnitude in the change of the amenity is not outside the range observed in the total sample. If so, the prediction is out of sample and becomes less reliable the further out of sample the change.

for those means) is predicted to sell for \$110,466. If the water clarity at the lake, where the property is located (a small part of the overall market) were increased from 3.9 to 6.3 m (the highest observation for this data), this property would be predicted to sell for \$115,133. The net benefit for this house is \$4,667. The total net benefit would be the sum of the net benefits estimated for each house located on this particular lake.

If one relaxes the assumption of costless moving on the part of renters, then the computation above represents an upper bound on the welfare change associated with the amenity change (i.e., overstate gains and understate losses). From the price change given in Eq. (7.12), one would need to subtract the transactions costs (TC) associated with each renter having to move to his or her new optimal location in order to compute the total change in welfare. For each property, net benefits are given by $P^1 - P^0 - \mathrm{TC}$. Palmquist (1992b) summarized the manner in which transactions and moving costs may be quantified and incorporated empirically into a hedonic analysis.

The above analysis assumes consumer utility is unchanged as a result of the change in the amenity (Δz_1) because renters move to a new house with the exact same bundle of characteristics as before the amenity change. This follows from the assumption that the change in the amenity does not affect the overall hedonic price function in the market but only affects the price of houses experiencing Δz_1. Thus, in optimizing utility, a house identical to the original dwelling before the amenity change will be chosen. If identical housing is not available and utility changes as a result of the move, then the correct welfare measure would require quantifying the difference in utility arising from the consumer no longer being at the optimal consumption bundle (see Bartik, 1988, for a complete description of this problem). While this computation is not possible, using information from only the hedonic price function, the change in rents less transactions costs would provide an upper bound on the resulting net benefits from an amenity change. The degree to which this bound is close to the true net benefits will depend on the degree of substitutability between the housing affected by the amenity change and the amenity level at the housing renters move to.

Table 7.3 summarizes the benefit estimates possible with the first-stage hedonic price function when the change in the amenity is localized. Note that the amenity change that occurs locally may be marginal or nonmarginal. The table highlights that simply computing the change in property values at locations will overstate the benefits of an amenity improvement and will understate the losses of an amenity decline when transactions costs are associated with moving (Bartik 1986).

When transaction costs prohibit moving, the net benefits of an amenity change cannot be exactly measured, but an upper bound can again be computed as the price differential for the property minus the transaction costs associated with moving. Consider a decline in an amenity. The landlord's capital loss as a result of the amenity change is exactly offset by the tenant's capital gain associated with paying a lower rent (recall that the tenant does not move). Thus, the price change is a transfer between owner and tenant, and no net benefit is associated with the change in rent. However, the reduction in rent only partially offsets the consumer's utility

Table 7.3 Benefits calculations with the hedonic price function

Computation	Appropriate situation
Cross-sectional estimators	
$\partial P/\partial z_i$	Net benefits of an ↑/↓ in the amenity by one-unit (a marginal change) in a localized context with zero transactions costs associated with moving
$P^1 - P^0$	Net benefits of an ↑/↓ in the amenity in a localized context with zero transactions costs associated with moving
$P^1 - P^0 - TC$	Net benefits of an ↑/↓ in the amenity in a localized context when transactions costs are positive, but do not prohibit moving. Household can relocate to an identical house as before the amenity change
$P^1 - P^0 - TC$	*Upper-bound* on net benefits of an ↑/↓ in an amenity in a localized context when (1) moving is possible, but households cannot relocate to an identical house as before the amenity change or (2) when it is *not* possible to move because transactions costs are prohibitively high
Quasi-experimental estimators	
ΔP	The treatment effect, commonly measured by $\partial P/\partial d_{jt}$ (see Eq. 7.10), is measure of net benefits if change is marginal and localized, but unlikely to equal net benefits for nonmarginal changes (even if localized)

loss. Transaction costs provide an upper bound on the remaining utility loss (after the change in rent is taken into account). Thus, the difference between the change in rent and transaction costs associated with moving will provide an upper bound on the total welfare change associated with the change in the amenity for a consumer (Table 7.3). Palmquist (1992b) presented a graphical exposition of this point.

In the case of *nonlocalized* amenities, an exact measure of welfare change cannot be computed with only information from the hedonic price function. This is because a nonlocalized change implies a shift in the overall supply, and the hedonic equilibrium itself will change. The new hedonic price function cannot be forecast using information from only the current price function. Bartik (1988) demonstrated that the change in hedonic rent at sites that are improved, as predicted by a hedonic price function estimated prior to a nonlocalized change in an amenity (call this measure ΔHP), is likely to be an upper bound for the net benefit of the amenity change. This bound is computed as given in Eq. (7.12) and is, therefore, quite easy to implement. However, ΔHP is only an upper bound under some conditions on the profit changes landlords may experience when allowed to make equilibrium adjustments to housing characteristics in response to the amenity change (Bartik 1988). These profit changes are likely to be in a range such that ΔHP will be an upper bound only if no transactions costs are associated with moving, which is likely implausible. If there are significant transactions costs, this measure is less likely to be an upper bound. Unfortunately, one cannot know if ΔHP is an upper bound, and empirical tests have not determined how likely it is to be an upper bound.

Note also, the computation of net benefits discussed above is directly linked to our understanding of what quasi-experimental designs measure (see Sect. 7.2.3.2). Recall, quasi-experimental estimates rely on an exogenous shock to the market—usually a nonmarginal shock—to identify changes in prices due to a change in an amenity. The treatment effect, or change in price due to a change in the amenity, is measured by β_0 in Eq. (7.10), and it is only equal to the underlying WTP for a change in the amenity under certain conditions. As Kuminoff and Pope (2014) showed, the shock to the environmental amenity must be uncorrelated with the initial conditions (i.e., the initial housing *and* environmental conditions) and changes in the housing stock over time. However, this may be problematic in practice because many environmental policies are by design correlated with initial environmental conditions. Further, it is hard to ensure that behavioral responses to changes in environmental amenities are not significant (e.g., a change in the composition of housing postenvironmental improvements). This last consideration can be alleviated by focusing the analysis on a relatively narrow time window around a particular shock, but data limitations may not always allow this either.

In general, the measure of ΔP (see Table 7.3) from a quasi-experiment will not equal net benefits when ΔP is measuring differences across two different hedonic equilibria (one prepolicy equilibrium and one postpolicy equilibrium) except in the limiting, unrealistic case when demand for an amenity is perfectly elastic.[22] Kuminoff and Pope (2014) also demonstrated that the bias in the measured ΔP (a "capitalization rate") will generally be in an unknown amount and direction, and need not be bounded by ex ante or ex post willingness to pay. Finally, even if the exogenous shock is randomized in a way that satisfies the conditions needed to identify marginal willingness to pay in the postpolicy equilibrium, caution is still needed in using postpolicy marginal willingness to pay to assess the benefits of a policy. For example, a postpolicy marginal willingness to pay of zero for a further improvement in environmental quality could still mean that the policy increased welfare substantially because prepolicy marginal willingness to pay was high, or hardly at all because prepolicy marginal willingness to pay was low (Kuminoff and Pope 2014).

7.4 Demand Analysis Within the Hedonic Framework

To estimate uncompensated demands for the *characteristics* of the differentiated good, information on the quantities of characteristics purchased and the marginal implicit prices of those characteristics from a hedonic analysis are combined with information on the socioeconomic characteristics of the purchasers. Estimating the demand for characteristics of a differentiated product using hedonic prices is often referred to as a second-stage analysis (estimating the hedonic price function was described as the first-stage in the previous section). While Rosen (1974) provided the framework for this method, it continues to be refined today. This section

[22]See also Klaiber and Smith (2013) who employed a simulation strategy to further explore the relationship between capitalization rates and willingness to pay for amenity changes.

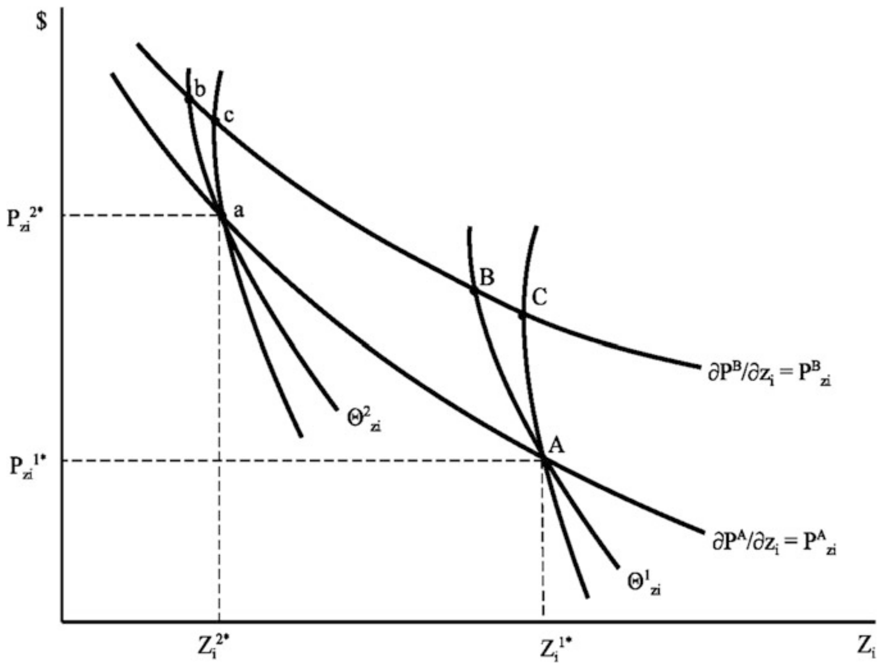


Fig. 7.2 Demand identification with the hedonic price function

provides an overview of second-stage methods. See Palmquist (2006) for a more technical discussion of the method.

For welfare measurement, researchers are typically interested in compensated demands, which indicate the willingness to pay for z_i, holding the levels of all other goods/characteristics and utility constant. Recall that the bid function for each individual, θ^j, is a function describing the maximum WTP for a specific bundle of characteristics, holding utility and income constant. Because bid functions change directly with income (i.e., $\partial\theta/\partial y = 1$), the marginal bid function for characteristic z_i, θ_{zi}, only depends on $\underline{z}$, utility, and exogenous consumer characteristics, α^j (Palmquist 2006). As such, the marginal bid function is equivalent to an inverse compensated demand function for characteristic z_i.[23] It describes the change in WTP for z_i as the level of z_i changes, holding utility and all other characteristics and goods constant.

Figure 7.2 illustrates the marginal bid function for two individuals, θ^1_{zi} and θ^2_{zi}. For the moment, consider only the marginal bid function for Consumer 1 as

[23]Note: The marginal bid function is not equivalent to a marginal rate of substitution function given by $(\partial U/\partial z_i)/(\partial U/\partial x)$, as had been suggested in the early literature. It is true that, at the *optimal level of consumption*, the marginal rate of substitution is equal to the marginal bid and the marginal price $(\partial U/\partial z_i)/(\partial U/\partial x) = \partial P/\partial z_i = \partial\theta/\partial z_i$, as discussed in Sect. 7.1.

depicted by the heavy dark line passing through points AB and the associated implicit price function P^A_{zi}. We can see that the marginal bid function is equal to the implicit price function at the optimal level of consumption for Consumer 1, z_i^{1*} (Point A). This point corresponds in Fig. 7.1 to the optimal consumption level where the bid function is tangent to the hedonic price function.

Theoretically correct measures of welfare change for a characteristic of a differentiated good are no different than was described for a quantity change in an amenity in Chap. 2. The literature typically makes the distinction between measures of welfare change when consumers face a price change (compensating and equivalent variation) and when consumers face an exogenous quantity change (compensating and equivalent surplus). These are analogous measures (see Chap. 2). In the case of an exogenous change for consumers who cannot move, compensating and equivalent surplus would be the appropriate measures. In the case of an exogenous change in z_i in which moving is possible, compensating and equivalent variation would be more appropriate. However, these measures will only measure the benefits to homeowners. If the change in the amenity affects consumers in other ways (say, increases recreation benefits to consumers who do not own nearby property but visit the park after it is improved), these benefits will not be captured using the measures discussed in this chapter.

Here, the focus is on compensating or equivalent surplus, as one typically assumes households cannot move in response to a change and thus computes the households' willingness to pay as a lower bound to the true welfare change (see Sect. 7.3). Measures of compensating or equivalent surplus are simply computed by integrating under the compensated inverse demand or the marginal bid function:

$$W(\Delta z_i) = \int_{z_{10}}^{z_{11}} \frac{\partial \theta(z_i; \underline{z}, U^j)}{\partial z_i} \mathrm{d}z_i, \tag{7.13}$$

where $W(\Delta z_1)$ is compensating or equivalent surplus for a change in z_i depending on whether U^j is equal to the initial level of utility or the level of utility after the amenity change, respectively. This measure of welfare change is appropriate for the situation in which consumers receive an exogenous quantity change in an amenity, and consumers are compensated in terms of the numeraire (holding z constant). Palmquist (2006) discussed welfare measures when other types of compensation are used to return consumers to their original (or new) level of utility.

Of course, estimating the marginal bid function directly is not possible because the marginal bid depends on unobserved utility. However, several alternatives allow us to recover the information necessary to compute welfare measures.

The first approach follows our discussion to this point and models the consumer's choice among a smooth continuum of housing bundles in which all possible combinations of characteristics are possible. In this approach, a second-stage analysis is used to recover Hicksian uncompensated demand if data from multiple markets are available.

The second approach relies on random utility modeling, which refocuses the modeling exercise to describe how a consumer makes a discrete choice to purchase a single house from a set of alternatives that he or she is considering. In this approach, utility parameters are directly estimated and welfare analysis follows directly. Each method is described in turn next.

7.4.1 Uncompensated Demands for Characteristics

Uncompensated demand functions can be derived analytically in a manner analogous to a homogenous goods market if the hedonic price function is linear so that the marginal prices are constant. However, researchers often expect the hedonic price function to be nonlinear and the marginal prices to be nonconstant. In this instance, the budget constraint is no longer an exogenous constraint; the marginal price of any characteristic is simultaneously determined by the choice of the amount of that characteristic to purchase. Traditional optimization methods are inappropriate in this case.

Uncompensated demands with nonconstant marginal prices can be derived, however, if we linearize the budget constraint around the optimal consumption bundle of the consumer. Palmquist (1988) demonstrated that utility maximization with this constraint set yields the same choices as would be observed with the original, nonlinear budget constraint. The linear budget constraint is of the form

$$Y_a^j = x + \sum_{i=1}^{n} P_i^* z_i, \tag{7.14}$$

where Y_a^j is Consumer j's income adjusted in the following manner:

$$Y_a^j = y - P(Z*) + \sum_{i=1}^{n} p_i^* z_i^*. \tag{7.15}$$

The linearized budget constraint in (7.14) is exogenous as the implicit prices, p_i^*, faced by Consumer j are held constant at the level associated with this consumer's actual purchase, z_i^*. Income must be adjusted as shown in Eq. (7.15) because a linear budget constraint will imply a larger consumption set than is affordable for the consumer with income y facing a budget constraint in which prices for z_i are decreasing as amounts of z_i are increased.

The linearized budget constraint in (7.14) may be used in conjunction with the first-order condition for utility maximization (given in Eq. (7.3)) to solve for the inverse uncompensated demand in a manner analogous to that for homogeneous goods:

$$P_1^j = f\left(z_1, z_2, \ldots z_n, x, Y_a, \alpha^j\right). \tag{7.16}$$

This demand function can be estimated using quantities of characteristics chosen, the implicit prices of those characteristics, adjusted income, and other socioeconomic characteristics assumed to affect demand.

Consumer surplus estimates for a change in z_i may be computed by integrating vertically under the inverse demand curve for z_i between the initial and the new level of z_i (Parsons 1986). For example, suppose the following simple linear demand for z_1 is estimated (note, this is not an inverse demand function):

$$z_1 = \alpha + \beta_1 p_1 + \beta_2 p_2 + \delta y, \tag{7.17}$$

where p_2 is the implicit price of a substitute/complement for z_1 and y is adjusted income. The integral of the inverse demand between the initial level z_{10} and new level z_{11} is

$$\int_{z_{10}}^{z_{11}} \frac{1}{\beta_1}(z_1 - k)\mathrm{d}z_1 = \left[\frac{1}{\beta_1} z_1^2 - \frac{k}{\beta_1} z_1\right]\Bigg|_{z_{10}}^{z_{11}}, \tag{7.18}$$

where k is a constant equal to $\alpha + \beta_2 z_2 + \delta y$ evaluated at some appropriate level, such as the mean observed in the data.

The demand function estimated should include relevant characteristics that are substitutes or complements to the characteristic of interest. Choosing the appropriate set of substitutes and complements may be somewhat ad hoc because no theoretical guidance defines exactly what is a substitute or complement for any one characteristic. A reasonable approach is to assume that the major characteristics of a property (such as house, lot size, and a few key neighborhood characteristics) are likely substitutes and complements to include. Boyle et al. (1999) estimated the following uncompensated demand for water clarity (Q_{wc}):

$$Q_{\mathrm{WC}} = f\left(P_{\mathrm{WC}}, P_{\mathrm{SQFT}}, P_{\mathrm{FRONT}}, Y_{\mathrm{adj}}, \Gamma\right), \tag{7.19}$$

and computed consumer surplus for changes in water clarity as described in Eqs. (7.17) and (7.18). In Eq. (7.19), Y is income as adjusted in Eq. (7.15) and P_{WC}, P_{SQFT}, and P_{FRONT} are the marginal implicit prices of water clarity, square feet of living space, and lake frontage of the property, respectively. They chose two major characteristics of the property: size of the house and the amount of lake frontage. Results indicated that both characteristics were substitutes for water clarity as the price coefficients for house size and lake frontage were positive and significant.

In addition to income, the researcher must determine the other relevant factors expected to shift demand, represented by the vector Γ in Eq. (7.19). As with any demand function, typical factors might include the age and education of the purchaser. For property markets, factors such as the number of children under 18 years old would also be expected to be important demand shifters. Other factors specific to each application might be included. For example, in the Maine lakes example, Boyle et al. (1999) included a dummy variable indicating whether or not the owner had friends who had bought property at the lake prior to his or her own purchase of property.

After the set of substitutes, complements, and demand shifters are chosen, the researcher must still specify the function to be estimated. A common functional form for demand is the semilog functional form. Boyle et al. (1999) estimated a linear, semilog, and Cobb–Douglass specification for Eq. (7.19), preferring the nonlinear specifications. Based on the semilog specification, the estimated own-price elasticity for water clarity was −0.158, and the estimated surplus for an improvement in lake water clarity from 3.78 to 5.15 m was \$3,765. The estimated surplus for the same change in quality was \$12,870 using linear specifications and \$3,677 using Cobb–Douglas specifications. As might be expected, the results are sensitive to the choice of functional form. A careful researcher investigates how sensitive their results are to choice of functional form (and specification of variables included). If they have a preferred equation for reporting, reasons for this preference should be explained.

Consumer surplus measures can be computed using estimates of the parameters of the uncompensated demand. However, consumer surplus is only an approximation of the theoretically correct measures of welfare change (see Chap. 2, and Freeman et al., 2014). Two approaches may be used to recover compensating or equivalent variation when uncompensated demands are estimated. In the first, utility is specified and a system of uncompensated demands are derived analytically and estimated. Given estimates of the demand parameters, duality may be used to analytically recover estimates of compensating or equivalent variation by solving for indirect utility or expenditure functions. This approach was taken by Palmquist and Israngkura (1999) and Parsons (1986).

Alternatively, one can estimate a single demand for a characteristic of interest and use differential equation methods to recover an associated utility or indirect utility function. Hausman (1981) and Vartia (1983) demonstrated how this may be accomplished when prices are changing. Palmquist (2006) demonstrated this method for quantity changes to solve for the bid function.

Consider the inverse uncompensated demand in Eq. (7.16). Recognizing that the marginal implicit price is equal to the marginal bid function at the optimal level of consumption, we can substitute $\partial\theta/\partial z_i$ for the left side of Eq. (7.16). Also, because our numeraire x is identically equal to $Y - \theta$, we can substitute for x in the right-hand side, and Eq. (7.16) becomes a differential equation in θ. Solving this differential equation for the bid function allows us to compute welfare measures as follows:

$$W(\Delta z_i) = \int_{z_i^0}^{z_i^1} \frac{\partial\theta(z_i, \underline{z}, U^j)}{\partial z_i} \mathrm{d}z_i = \int_{z_i^0}^{z_i^1} \mathrm{MRS}\left(z_i, \underline{z}, x\left(z_i, \underline{z}, U^j\right)\right), \tag{7.20}$$

which is equal to compensating surplus if utility is held constant at the original level of utility, and equal to equivalent surplus if utility is held constant at the level of utility after the change (note: income is held constant).[24]

7.4.1.1 Identification of Uncompensated Demand

Recall that each consumer will choose an optimal level of each characteristic by finding the level of consumption at which the marginal bid for z_i equals the implicit price of z_i. By estimating the parameters of $P(z)$ and computing $\partial P(\underline{z})/\partial z$, one can observe one point on each consumer's uncompensated demand as well as his or her compensated demand function or marginal bid function. (The term "marginal bid function," rather than inverse compensated demand function, is used for consistency with past discussions.) However, no other information on the shape of the bid function is observed. Consider Fig. 7.2 again, which shows the marginal bid functions for two consumers and the marginal price function P_{zi}^{A} (ignore for the moment the second marginal price function, P_{zi}^{B}). For reasons discussed in Sect. 7.2.1, the hedonic price function is likely to be nonlinear, so the marginal price function is likely to be nonconstant. In Fig. 7.2 the marginal price of z_i is shown to decrease as more of z_i is consumed.

Figure 7.2 indicates that each marginal price reveals only one point on each consumer's marginal bid function, so one cannot make inferences about the shape of the marginal bid function using only information on marginal prices. By observing P_{zi}, it is known that one point on the marginal bid function of Consumer 1 is $\{P_{zi}^{1*}, z_i^{1*}\}$; however, it cannot be determined if the demand function is shaped as shown by the solid line or the dashed line. Indeed, an infinite number of possible functions could pass through the point $\{P_{zi}^{1*}, z_i^{1*}\}$. This is the standard problem of identifying any demand function: Additional information besides the market-clearing prices for each property in a market is needed to identify demand.

One approach to identifying demand is to use information from separate markets, such as different cities that are geographically distinct. In this approach, one assumes that individuals in one market do not consider housing in the other markets as substitutes when making purchase decisions. Thus, the hedonic price function in each market is distinct and arises from interactions of only those buyers and sellers

[24] An alternative method for recovering the information necessary to compute compensating and equivalent variation is to directly specify the functional form for household utility and derive the functions to be estimated (see Quigley 1982). In other words, the exact specification of the regression is determined by the analytical solution to the utility maximization problem. What function is estimated depends on the goal of the research. For estimating welfare changes, Cropper et al. (1993) and Chattopadhyay (1999) estimated the equilibrium condition given in Eq. (7.3) in which marginal prices (computed from the first-stage hedonic regression) are equal to the marginal rate of substitution function. While this function may not be used directly for computing welfare changes, it recovers utility parameters that can then be used to compute estimates of welfare using duality results.

located in that market. Individuals with any given vector of characteristics are assumed to have preferences over attributes that are identical across markets. However, because of differences in supply, or perhaps because of the distribution of socioeconomic characteristics among individuals within a market, the equilibrium hedonic price functions will vary across markets, so similar people will be observed making different housing choices across the different markets. If this is the case, estimating separate hedonic price functions in each market will identify demand. This point is illustrated in Fig. 7.2, where P^B_{zi} represents the marginal price function from a separate market. Given this additional information, we can now determine if it is Point *B* or *C* that is the optimal choice for that type of consumer and whether the marginal bid function is represented by the dashed or the solid line.

Thus, to estimate an inverse uncompensated demand (such as in Eq. (7.16)) or a system of demands, hedonic price functions are estimated for multiple markets. The marginal prices from each market are pooled with information on property owners' demographics and characteristics of their properties to estimate demand. For example, to estimate the demand given in Eq. (7.19), Boyle et al. (1999) specified a hedonic price function like that given in Eq. (7.1) for four markets in Maine. Each market was distinguished by distances between each other, by unique regional characteristics, and by different real estate multiple listing service regions. The estimated marginal implicit prices along with corresponding quantities purchased were pooled with demographic information from a survey of property owners.

The use of multiple markets to identify demand for housing attributes has also been used by Palmquist (1984), Parsons (1986), Bartik (1987), Cheshire and Sheppard (1998), Palmquist and Israngkura (1999), and Zabel and Kiel (2000), with the latter two studies focused on environmental amenities. The number of markets used to identify demand has varied from two (Cheshire and Sheppard 1998; Bartik 1987) to seven (Palmquist 1984; Parsons 1986) to 13 (Palmquist and Israngkura 1999). There is no established correct minimum number of individual markets required. Of course, what must be established is that the hedonic price functions do vary across markets; thus, researchers must provide evidence that statistically significant differences in the implicit prices across markets exist. The issues associated with determining appropriate assumptions for market segmentation are the same here as are discussed in Sect. 7.2.1.2.

Kuminoff and Pope (2012) proposed an alternative strategy within a quasi-experimental setting. Specifically, the authors show that if a natural experiment meets the requirements to be a good instrument and provides a large shock so that the hedonic equilibrium shifts, then this known, exogenous shift in the price function may be used to identify the inverse uncompensated demands. This point can be illustrated by reviewing Fig. 7.2, which highlights how identification could be achieved with information from separate markets. In a quasi-experimental setting, one could estimate a single-period hedonic price function preshock (akin to P^A_{zi} in Fig. 7.2) and a separate single-period hedonic price function postshock (akin to P^B_{zi} in Fig. 7.2) to provide identification for the inverse bid functions.

7.4.1.2 Endogeneity of Prices and Income in Demand

In addition to identifying demand, an important econometric issue that must be taken into account is the possible endogeneity of implicit prices and income. Recall that for any functional form of the hedonic price function other than linear, as given in Eq. (7.6), implicit prices may be nonconstant. If prices are nonconstant, then they are endogenous because consumers simultaneously choose the marginal price they pay per unit of the characteristic when they choose the level of the characteristic. Also, in the case of nonconstant implicit prices, we linearized the budget constraint to analytically derive the demand function, which involved adjusting income by the implicit prices (Eqs. (7.14) and (7.15)). When including adjusted income in the demand specification, it too will be endogenous because the adjusted income relies on the nonconstant implicit prices.

For instance, the hedonic price function estimated by Boyle et al. (1999) in Eq. (7.1) indicates that the marginal price for water clarity and square feet of living space are nonconstant. Thus, by choosing the quantity of living area, for instance, the purchasers are also choosing the marginal price per foot of living area they must pay. The marginal price of living area and the choice of square feet of living area are both endogenous.

Endogeneity is typically handled by instrumental variable techniques. In this method, each price that is endogenous and adjusted income is first regressed on a set of exogenous explanatory variables. These exogenous variables or instruments should be related to the endogenous variable of interest but exogenous to the system. The instruments must be (1) correlated with the regressors, (2) uncorrelated with the error term, and (3) of full rank (that is, adds new information). The resulting parameter estimates from the first-stage regression are used to predict values for the endogenous prices and income. The predicted prices and predicted income are then used to estimate the demand equation (see Greene, 2011, for a standard discussion of two-stage least squares).

Choosing the proper instruments for the endogenous variables can be a difficult task. For the demand given in Eq. (7.19), Boyle et al. (1999) developed nine instruments for the marginal prices that described factors such as number of lakes in a market area, distance of the market to Boston, a time trend, and local economic and demographic conditions (e.g., an index of real estate activity in the area and the percentage of the current population that is employed). For adjusted income, the instruments were socioeconomic characteristics of the property purchasers (as reported in a survey of the owners) and included age and gender of the owner, number of children under 18-years old, educational level, retirement status, and the number of people in the household. Also included were the squared terms for age, the number of children under 18-years old, and the number of people in the household.

Palmquist (1984) estimated the demand for four housing characteristics; he used exogenous socioeconomic characteristics (age, marital status, number of dependents, race, and their square where appropriate), and he used a set of dummy variables for each of his seven urban areas to instrument for the nonconstant implicit prices and adjusted income.

Cheshire and Sheppard (1998) used spatially lagged variables as their instruments. The authors developed a spatial relationship between each house in their data set and each next-closest house.[25] For each observation, the characteristic prices and income levels associated with the two houses closest to the house in question were used as instruments. The authors argued that these variables clearly meet requirements (1) and (3) for instrumental variables that were discussed earlier and are likely to meet requirement (2) because there was enough variability in neighborhood housing characteristics (and individual characteristics within a neighborhood). Palmquist (2006) made the point that because this approach is not unidirectional (i.e., the spatial lags are not unidirectional in the same way time lags are), these instruments are invalid. To alleviate this problem, one might combine time and the spatial lags to create an instrument for the implicit price that is the closest house sold prior to the current house.

A common issue is that instruments do not explain the variation in the endogenous variables very well. In finite samples, weak instruments result in biased estimates of the demand parameters, just as in the case of using simple, ordinary least squares (Bound et al. 1995). Thus, with weak instruments, one may be accepting higher variance of the parameter estimates without reducing the bias. Also, if instruments are weak and the loss of efficiency is large, common tests for endogeneity (Hausman 1978) are likely to have weak power (Nakamura and Nakamura 1998). In addition, a loss of relevance can occur when the instruments chosen for the endogenous regressors do not adequately capture the type of variation in the regressor. For these reasons, Nakamura and Nakamura rejected the "always instrument" policy for endogeneity—especially for variables that are not of primary interest to the application or when it is not clear that the potential instruments are truly exogenous. While there is no clear guidance for what are weak instruments, R^2 values for the auxiliary equations of less than 0.2 caused concern for Nakamura and Nakamura.

7.4.2 Discrete-Choice Models of Housing Demand

A different approach to identifying demand parameters recasts consumer decisions in a random utility framework.[26] In a random utility model, consumers are assumed to make a discrete choice between house bundles, rather than a continuous choice over attribute levels as in the typical hedonic framework. In the random utility

[25] Closeness is defined by both geographic proximity and similarity of the structures.

[26] See Klaiber and Kuminoff (2014) and Kuminoff et al. (2013) for detailed reviews. Phaneuf and Requate (in press) also provided a general overview. Note, the model discussed here in the context of housing choice is conceptually and empirically equivalent to the random utility model described for travel cost models in Chap. 6.

model framework, the consumer knows his or her utility for all possible choices, but utility is measured with error because not all factors that influence choices are accessible to researchers. Thus, utility (U) is assumed to be the sum of a systematic portion (V) and a random component. This can be written as follows for a consumer, j, who is observed choosing a house, k:

$$U^j\left(x^k, \underline{z}^k; \alpha^j\right) = V^j\left(x^k, \underline{z}^k; \alpha^j\right) + \varepsilon_{jk}, \tag{7.21}$$

where $U(\cdot)$ is the true, but unobservable utility as defined in Eq. (7.2), and ε is the error term introduced because the researcher cannot observe all relevant aspects of the individual. The individual maximizes utility by selecting the house that yields the highest level of utility from among the set of all possible alternatives, A. The probability that Consumer j selects House k is given by

$$\begin{aligned} &\Pr(k|A) = \Pr(U_k \geq U_m) = \Pr\left\{V\left(x^k, \underline{z}^k; \alpha^j\right) + \varepsilon_{jk} \geq V\left(x^m, \underline{z}^m; \alpha^j\right) + \varepsilon_{jm}\right\}, \\ &\forall \;\; k, m \ni A \text{ and } k \neq m. \end{aligned} \tag{7.22}$$

The assumption about the distribution of the random error term implies which probabilistic model is estimated. A common assumption is that the error terms are independently and identically distributed following a Type 1 extreme value distribution, leading to the well-known conditional logit model. To relax the independence of irrelevant alternatives assumption inherent in the conditional logit model, a nested logit model or a random parameters logit may be used.

Welfare measurement for a marginal change in z_i is given by $(\partial V/\partial z_i)/(\partial V/\partial x)$, and for a nonmarginal change, the willingness to pay, C, for a change in an amenity of interest, say z_1, is given by $V\left(x^0, z_1^{k0}, z_2^k, \ldots z_n^k; \alpha^j\right) = V\left(x^0 - C, z_1^{k1}, z_2^k, \ldots z_n^k; \alpha^j\right)$. Note that in this formulation, the price paid for the house is held constant at the purchase price. Substitution of the budget constraint, $x^0 = y^j - P(\underline{Z}^k)$, makes this clear. In this sense, the welfare measures from the random utility model approach rely on the current distribution of prices. Further, because the random utility model in itself does not model the market equilibrium, it cannot be used to infer how a nonlocalized change in an amenity would affect the equilibrium price schedule. Thus, while the random utility model can directly compute welfare measures based on the current price vector, additional steps are needed to estimate general equilibrium welfare estimates for a nonlocalized amenity change (e.g., air quality changes across a city).

Sorting models extend the random utility model to address general equilibrium responses to a nonlocalized change in an amenity. Environmental applications of sorting models shift the focus of the analysis from the choice of a house to the choice of a neighborhood and add an equilibrium condition equating demand and supply for each neighborhood. Environmental amenities in these models vary across neighborhoods, and housing supply is typically assumed to be fixed. Aggregate demand is equilibrated to fixed supply by the equilibrium price vector for the n neighborhoods in the model,$\{P_1, \ldots P_n\}$.

Demand in a sorting model is defined by a household's expected demand for each location. Equation (7.22) can be recast to represent this demand by letting k denote neighborhoods rather than houses and having P and z be indexes of these variables computed by the researcher to vary by neighborhood (see Klaiber and Phaneuf, 2010, for a discussion). Substituting the budget constraint into Eq. (7.22) and summing over individual demands yields the expected aggregate demand for neighborhood k:

$$\mathrm{AD}_k(P, z, \alpha) = \sum_j \mathrm{Pr}_{jk}\left(P, z, \alpha^j\right), \tag{7.23}$$

where P and z are the vector of prices and housing characteristics across the entire landscape (i.e., for all k neighborhoods). Writing Eq. (7.23) in share form gives the proportion of market participants who are expected to select neighborhood k, given a set of prices and attributes across the landscape. We then define an equilibrium condition, which requires that the share of demand for any particular neighborhood, k, must be equal to the proportion of total housing in the market that is available in neighborhood k:

$$s_k^D(P, z, \alpha) = \frac{1}{J} AD_k(P, z, \alpha) = s_k^S. \tag{7.24}$$

The market-clearing condition in Eq. (7.24) is important for welfare analysis. Given a shock to z (say, an improvement in air quality), Eq. (7.24) can be solved for a new set of market-clearing prices postshock. As illustrated below, this allows computation of welfare effects from a change in z that do not require prices to be held at their original, preshock levels as traditional random utility models require.

In addition, a particularly nice feature of these models is the straightforward manner in which they allow the researcher to incorporate unobserved neighborhood attributes. To see this and how the mechanics of estimation proceed, rewrite utility in Eq. (7.21), focusing on the choice of a neighborhood:

$$U^j\left(P^k, \underline{z}^k, q^k, \zeta^k; \alpha^j\right) = \alpha P^k + \beta Z^k + \delta q^k + \gamma \alpha^j q^k + \zeta^k + \varepsilon_{jk}, \tag{7.25}$$

where utility now depends on an index of housing price in neighborhood k, P^k, housing characteristics in neighborhood k, Z^k, and an environmental attribute of interest in neighborhood k, q^k. Equation (7.25) allows for heterogeneity in preferences for the environmental amenity through the term $\gamma \alpha^j q^k$. Further, neighborhood unobservables are explicitly incorporated in the model through ζ^k, a measure of neighborhood attributes that is observable to the decision-maker but unobserved by the researcher.

Equation (7.25) may be rewritten in a simplified form that makes clear the role of attributes that vary by neighborhood and those that vary by individual:

$$U^j\left(P^k, \underline{z}^k, q^k, \zeta^k; \alpha^j\right) = \mu^k + \gamma\alpha^j q^k + \varepsilon_{jk}, \tag{7.26}$$

where

$$\mu^k = \eta P^k + \beta Z^k + \delta q^k + \zeta^k. \tag{7.27}$$

In Eq. (7.26), the choice of the kth neighborhood depends on the population mean utility for neighborhood k (average tastes), and $\gamma\alpha^j q^k$ differentiates utility away from the mean for households based on observable characteristics, α^j.

Estimation now proceeds in two steps. In the first step, Eq. (7.26) is estimated by assuming the error in Eq. (7.21) is distributed Type I, extreme value leading to the conditional logit model. The conditional logit model has the convenient property that the estimated demand share for any particular neighborhood will equal the observed share, and these shares will sum to one across all neighborhoods. In other words, the conditional logit model imposes the equilibrium condition on the data as an artifact of the model's statistical properties. Given the assumed utility in Eq. (7.26), this can be written formally as

$$s_k^D \equiv \frac{1}{J}\sum_j^J \left(\frac{\exp(\mu^k + \gamma\alpha^j q^k)}{\sum_m \exp(\mu^m + \gamma\alpha^j q^m)}\right) = s_k^S. \tag{7.28}$$

The maximum likelihood estimates of μ^k and γ assure Eq. (7.28) holds.

In the second step, the estimated μ^k are then used to estimate Eq. (7.27), which decomposes the μ^k into observable (P^k, Z^k, q^k) and unobservable (ζ^k) components. In this stage, endogeneity concerns surround the correlation between P^k and ζ^k. Given that the estimation of Eq. (7.27) is linear and estimated by ordinary least squares, standard econometric tools (instrumental variables) are available to address this endogeneity.

Given the full set of parameter estimates, general equilibrium welfare effects of a policy change can be evaluated. With a change in q^k, Eq. (7.28) is used to solve for the new price vector. Welfare measures that incorporate both changes in q and changes in prices are then available to compute based on the assumed form for utility. Klaiber and Phaneuf (2010) implemented this framework to explore the value of open space amenities. See also Tra (2010, 2013) for applications valuing air quality.

It is worth repeating that strategies for identifying unbiased coefficient estimates must be considered carefully in the random utility model framework as well. Even though random utility models and sorting models operate somewhat differently, they still require careful consideration of endogenous variables due to simultaneous determination or omitted variable bias just as with multimarket approaches. In addition, random utility models and sorting models require careful consideration about the definition of choice sets and the unit of analysis (a neighborhood).

7.5 Labor Markets and the Value of a Statistical Life

In addition to housing, a second important application of the hedonic method in nonmarket valuation is labor markets (Rosen 1974). To see how hedonic modeling applies within a labor market context, reconsider Fig. 7.1. In a labor market context, the hedonic price function is now a hedonic wage equation, and it is an envelope of individual firm's offer functions and workers' indifference curves. For example, Fig. 7.1 could be the hedonic wage gradient for a specific job characteristic, such as a measure of the on-the-job risk of injury. The wage trade-offs that workers are willing to make to accept a higher risk of on-the-job injury, holding utility constant, are represented by Φ. Higher contour levels of Φ would represent higher levels of utility. Similarly, θ are now iso-profit curves of the firm, representing the trade-offs the firm can make between higher wages and expenditures to reduce the risk of on-the-job injury, holding profit constant. Lower contour levels of θ represent higher levels of profits (see Bockstael and McConnell, 2007, for an excellent technical review).

At first blush, hedonic wage models may not seem directly related to environmental valuation. However, as Cameron (2010) highlighted, they continue to be the subject of vigorous academic and policy debates because they are a primary way in which one measures the benefits of federal environmental (and other) regulations that reduce human mortality risks. Specifically, hedonic models are used to estimate the trade-offs workers are willing to make between wages and the risk of death on the job. These dollar trade-offs can then be used to compute a measure referred to as the "value of a statistical life" or VSL. The VSL is not the value of saving a particular individual (an "identified life"), but the value society places on a reduction in the probability of one death among them (see Viscusi 1992, 1993).

The easiest way to understand the VSL is through an example. Consider the following simple wage hedonic:

$$\text{wage}_k = \alpha + \beta_r \text{risk}_k + \sum_{n=1}^{N} \lambda_n X_{kn} + \sum_{m=1}^{M} \gamma_m D_{km} + \varepsilon_k, \tag{7.29}$$

in which the wage of the kth worker is estimated to be a function of the risk of death on the job (risk_k); N variables describe human capital and demographic characteristics of the worker (X_{kn}), such as age and education; and M job characteristics (D_{km}) other than the risk of death, such as whether or not supervisory activities are associated with the job.[27] In a linear specification such as (7.29), the implicit wage for risk, $\partial w/\partial r = \beta_r$, is the additional wages a worker would require to assume an additional increment of risk of death on the job. By normalizing over the risk

[27]The compensating wage equation describes the equilibrium wage as a function of the factors affecting the wage negotiation. This negotiation will be in regard to factors such as job duties and working conditions, as well as characteristics of the workers themselves that affect expected productivity.

increment, the compensating wage differential for risk is converted to the value of a statistical life. For example, suppose risk is measured in units of deaths per 10,000 workers, wages are hourly earnings, and a simple linear compensating wage equation is estimated as illustrated in (7.29). To compute the value of a statistical life, β_r is multiplied by 2,000 h per year to arrive at the annual compensation required for an additional unit of risk (assuming workers work 40 h per week for 50 weeks per year). The annual compensation is then multiplied by 10,000, the number of workers over which the risk is spread. In this example, an estimate of β_r equal to 0.35 would imply a VSL estimate of $7 million. Reviews of the early hedonic wage literature estimating the VSL are available in Mrozek and Taylor (2002), Viscusi and Aldy (2003), and Kochi et al. (2006).

In some environmental policies, particularly air quality policy, the benefits from mortality reductions can be the largest single component of the monetized benefits associated with the policy. For example, in the U.S. Environmental Protection Agency's (1997, 1999) retrospective (1970-1990) and prospective (1990-2010) analyses of the Clean Air Act, $4.8 million (1990 dollars) was the point estimate for the VSL used. This value was then aggregated over the number of lives saved to compute the monetized benefit of reduced mortality. In the retrospective analysis alone, the mean benefits of reduced mortality were estimated to be nearly $18 trillion (1990 dollars; U.S. Environmental Protection Agency 1997), just over 80% of the total benefits estimated to be associated with the Clean Air Act from 1970 to 1990.

From an empirical standpoint, many of the same issues considered for housing hedonic applications apply directly to labor markets. First, omitted variables and measurement error are as much a challenge in this literature, if not more so. Unobserved worker characteristics and job characteristics are likely correlated with job risks and wages and, therefore, bias the coefficient estimates for risk in an unknown direction (Black et al. 2003). While unobserved worker characteristics such as motivation or idiosyncratic job skills can be addressed within a panel-data framework (e.g., Kniesner et al. 2012), unobserved job characteristics are not.[28] Second, most studies rely on national average risk rates for broad occupational groups in broad industrial sectors. These measures are subject to considerable measurement error given their coarse resolution and the potential for dramatic differences among worksite safety practices even within the same industry (Lee and Taylor 2015).

Recently, quasi-experimental approaches have been used to address these concerns. Lee and Taylor (2015) provided the first quasi-experimental estimates

[28]Panel-data models that follow workers over time as they switch jobs (and thus, the risk of death on the job) do not control well for unobserved job characteristics because the models rely on job-changers to identify the marginal price of risk (i.e., β_r in Eq. (7.29)). To include individuals who do not change jobs in a panel-data framework, the researcher must rely on variation in workplace risks from year to year within the same job. Over relatively short periods of time, these variations are likely to be mostly random fluctuations in accidents. Over longer periods of time, changes in risk for any one specific job may be structural, but there is no reason to believe that there are not also changes in the correlated unobservables.

derived from a broad labor market setting. They employed federal workplace inspections that randomly select manufacturing plants for comprehensive safety inspections and use these inspections as an instrument for plant-level risks. The intuition behind the estimator is that prior to inspection, wages at riskier plants should be higher than those at less risky plants. Inspections provide an exogenous shock to the safety at plants that are inspected (an improvement), and wages are expected to fall (or rise less quickly) postinspection relative to uninspected plants whose safety remains unchanged. The estimator used to credibly identify wage-risk tradeoffs for manufacturing employees, and thus the VSL, is similar to the difference-in-differences estimator discussed in Sect. 7.2.3.2.[29]

A second important concern that has only partially been addressed within the hedonic framework is whether or not a VSL estimate derived from studies of fatal workplace risks is an appropriate value for reducing the risks that are regulated with environmental policy, such as the risk of premature death from exposure to poor air quality or cancer-causing agents (Cameron 2010; Scotton and Taylor 2011; Kochi and Taylor 2011). Deaths from cancers, illnesses, or heart attacks related to these environmental exposures may be associated with prolonged reductions in the quality of life prior to death or may involve more dread and pain as compared to workplace accidental deaths that are typically instantaneous. The author is unaware of research that estimates wage trade-offs workers make for on-the-job exposure to latent risks of death, such as exposures to cancer-causing agents.[30] This may be difficult to implement due to a lack of data and measurement difficulties but would be an important extension to the literature. In a related vein, a few studies have estimated the value of avoiding cancer-related illnesses within a housing hedonic context. Davis (2004) computed the implied estimate of a "VPL"—the value of (avoiding a case of) pediatric leukemia—to be \$5.6 million, and Gayer et al. (2000)

[29]Other quasi-experimental estimates of the VSL in nonlabor market settings also find lower VSL estimates. For example, Ashenfelter and Greenstone (2004) used changes in highway speed limits as a natural experiment. They first estimated the amount of saved travel time in states adopting increased speed limits and then estimated the impact of increased speed limits on rural interstate fatality rates. They reported an upper-bound point estimate for the VSL of \$2.2 million in 2011 dollars. León and Miguel (2013) also used a transportation setting to estimate the VSL, but within a developing country context. They used transportation mode choices (ferry, helicopter, hovercraft, and water taxi) and weather as an instrument for risk associated with each mode of travel to estimate a VSL of approximately \$600,000 for African travelers and approximately \$1 million for non-Africans.

[30]Scotton and Taylor (2011) and Kochi and Taylor (2011) attempted to address this issue by exploring the heterogeneity among workplace instantaneous deaths. Specifically, they note that nearly 20% of workplace deaths were due to homicide during the mid-1990s. Homicide risks have been shown to be viewed by individuals with more dread and fear than traditional job-related accidental risks, such as traffic and machinery accidents. Correspondingly, both studies find that homicide risks are compensated differently than traditional job-related accidental risks. While not conclusive, the authors argue that the results are suggestive that the VSLs derived from instantaneous workplace risks are not likely to be good estimates of the value of reducing risks associated with environmental exposures.

estimated the value of avoiding a cancer case to be $4.3 million to $5 million (all estimates are 2000 dollars).

A somewhat related concern—and one for which there is more research available—is whether or not the value of a statistical life derived from labor market studies based on working-age healthy adults is appropriate to extend to other populations at risk, such as the elderly or children. This is especially important for federal benefit-cost analysis of air quality regulations because much of the avoided mortality benefits of improved air quality accrues to the elderly population (for further discussion, see Eeckhoudt and Hammitt, 2001, Smith et al., 2004, Evans and Smith, 2006, and Aldy and Viscusi, 2007).

7.6 Summary

This chapter focused on the possibilities for nonmarket valuation using hedonic methods, primarily with property markets. Estimation of the hedonic price function alone (a first-stage analysis) is by far the most common hedonic application. Indeed, a search using a popular academic search engine using the keyword “hedonic” revealed more than 850 published works in the economics literature. The popularity of the hedonic method is not surprising given the relatively minimal data requirements and straightforward empirical implementation. In addition, the insights that may be drawn from a first-stage analysis are quite powerful: the value of a change in an environmental amenity that accrues to a population of residents in an area. But, it was also demonstrated that many details require careful consideration when estimating the hedonic price function. Current hedonic applications are expected to have intensive data and estimation procedures. Researchers must carefully consider their data sources, choices of the independent variables and their measurement, the functional form of the price function, the sample frame used, and the estimation procedures employed. While not all issues will require in-depth treatment, the case must be made that the researcher has identified without bias the implicit price they seek to measure.

Even though very many first-stage hedonic analyses have been conducted, this is still an exciting area of research. Advances in computing and explosive growth in freely available geospatial data have resulted in spatial modeling being fully integrated into empirical analyses. Quasi-experimental designs for evaluating the impact of changing environmental amenities on property prices have also grown rapidly in recent years, and this too will be an area for expanding research.

Beyond the first-stage estimation of the hedonic price function, much additional research can be conducted on estimating welfare changes within the hedonic framework. This is an important component of the hedonic nonmarket valuation literature because many policy changes are not localized in context, thus requiring demand analyses. Although welfare analysis in this context has been actively discussed since Rosen’s (1974) seminal article, it has remained relatively rare in implementation. More recently, however, sorting models have gained traction as an alternative to the data-intensive multiple-market approach to estimating demand. Here again, as rich data on properties *and* their owners are ever increasing in their availability, the scope for demand analyses can only increase.

Acknowledgements The author would like to thank Kevin Boyle, Tom Brown, Patty Champ, Nicolai Kuminoff, Jeffrey Zabel, and the participants in the nonmarket valuation workshop sponsored by the Rocky Mountain Research Station, April 2013, for helpful comments.

References

Aldy, J. E. & Viscusi, W. K. (2007). Age differences in the value of statistical life: Revealed preference evidence. Review of Environmental Economics and Policy, 1, 241-260.

Anselin, L. (1988). Spatial econometrics: Methods and models. Dordrecht, The Netherlands: Kluwer Academic.

Ashenfelter, O. & Greenstone, M. (2004). Using mandated speed limits to measure the value of a statistical life. Journal of Political Economy, 112, S226-S267.

Bajari, P., Fruehwirth, J. C., Kim, K. I. & Timmins, C. (2012). A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution. American Economic Review, 102, 1898-1926.

Bajari, P. & Kahn, M. E. (2005). Estimating housing demand with an application to explaining racial segregation in cities. Journal of Business & Economic Statistics, 23, 20-33.

Baranzini, A. & Ramirez, J. V. (2005). Paying for quietness: The impact of noise on Geneva rents. Urban Studies, 42, 633-646.

Baranzini, A. & Schaerer, C. (2011). A sight for sore eyes: Assessing the value of view and land use in the housing market. Journal of Housing Economics, 20, 191-199.

Bartik, T. J. (1986). Neighborhood revitalization's effects on tenants and the benefit-cost analysis of government neighborhood programs. Journal of Urban Economics, 19, 234-248.

Bartik, T. J. (1987). The estimation of demand parameters in hedonic price models. Journal of Political Economy 95:81-88.

Bartik, T. J. (1988). Measuring the benefits of amenity improvements in hedonic price models. Land Economics, 64, 172-183.

Bin, O., Crawford, T. W., Kruse, J. B. & Landry, C. E. (2008). Viewscapes and flood hazard: Coastal housing market response to amenities and risk. Land Economics, 84, 434-448.

Bin, O. & Landry, C. E. (2013). Changes in implicit flood risk premiums: Empirical evidence from the housing market. Journal of Environmental Economics and Management, 65, 361-376.

Black, D., Galdo, J. & Liu, L. (2003). How robust are hedonic wage estimates of the price of risk: The final report. Report submitted to U.S. Environmental Protection Agency, National Center for Environmental Economics, Washington, DC. http://yosemite.epa.gov/ee/epa/eerm.nsf/vwAN/EE-0483-04.pdf/$file/EE-0483-04.pdf.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. The Quarterly Journal of Economics, 114, 577-599.

Bockstael, N. E. & McConnell, K. E. (2007). Environmental and resource valuation with revealed preferences: A theoretical guide to empirical models. Vol. 7 of I. Bateman (Ed.), The economics of non-market goods and resources. Dordrecht, The Netherlands: Springer.

Boes, S. & Nüesch, S. (2011). Quasi-experimental evidence on the effect of aircraft noise on apartment rents. Journal of Urban Economics, 69, 196-204.

Bound, J., Jaeger, D. A. & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association, 90, 443-450.

Boyle, K., Lewis, L., Pope, J. & Zabel, J. (2012). Valuation in a bubble: Hedonic modeling pre- and post-housing market collapse. AERE Newsletter, 32 (2), 24-32.

Boyle, K. J., Poor, P. J. & Taylor, L. O. (1999). Estimating the demand for protecting freshwater lakes from eutrophication. American Journal of Agricultural Economics, 81, 1118-1122.

Brady, M. & Irwin, E. (2011). Accounting for spatial effects in economic models of land use: Recent developments and challenges ahead. Environmental and Resource Economics, 48, 487-509.

Brander, L. M. & Koetse, M. J. (2011). The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results. Journal of Environmental Management, 92, 2763-2773.

Buck, S., Aufhammer, M. & Sunding, D. (2014). The Value of Heterogeneous Property Rights and the Costs of Water Volatility. American Journal of Agricultural Economics, 96, 953–969.

Cameron, T. A. (2010). Euthanizing the value of a statistical life. Review of Environmental Economics and Policy, 4, 161-178.

Campbell, J. Y., Giglio, S. & Pathak, P. (2011). Forced sales and house prices. American Economic Review, 101, 2108-2131.

Carson, R. T. & Dastrup, S. R. (2013). After the fall: An ex post characterization of housing price declines across metropolitan areas. Contemporary Economic Policy, 31, 22-43.

Chattopadhyay, S. (1999). Estimating the demand for air quality: New evidence based on the Chicago housing market. Land Economics, 75, 22-38.

Chay, K. Y. & Greenstone, M. (2005). Does air quality matter? Evidence from the housing market. [Electronic version]. Journal of Political Economy, 113, 376-424.

Cheshire, P. & Sheppard, S. (1995). On the price of land and the value of amenities. Economica, 62, 247-267.

Cheshire, P. & Sheppard, S. (1998). Estimating the demand for housing, land, and neighbourhood characteristics. Oxford Bulletin of Economics and Statistics, 60, 357-382.

Cohen, J., Blinn, C. E., Boyle, K. J., Holmes, T. P. & Moeltner, K. (2014). Hedonic valuation with translating ameniities: Mountain pine beetles and host trees in the Colorado front range. Environmental and Resource Economics, 59, December.

Coulson, N. E & Zabel, J. E. (2013). What can we learn from hedonic models when housing markets are dominated by foreclosures? Annual Review of Resource Economics, 5, 261-279.

Cropper, M. L., Deck, L. B., Kishor, G. & McConnell, K. E. (1993). The estimation of consumer preferences for attributes: A comparison of hedonic and discrete choice approaches. The Review of Economics and Statistics, 75, 225-232.

Cropper, M. L., Deck, L. B. & McConnell, K. E. (1988). On the choice of functional form for hedonic price functions. The Review of Economics and Statistics 70, 668-675.

Davis, L. W. (2004). The effect of health risk on housing values: Evidence from a cancer cluster. American Economic Review, 94, 1693-1704.

Davis, L.W. (2011). The effect of power plants on local housing values and rents. Review of Economics and Statistics, 93, 1391-1402.

Eeckhoudt, L. R. & Hammitt J. K. (2001). Background risks and the value of a statistical life. Journal of Risk and Uncertainty, 23, 261-279.

Epple, D., Quintero, L. & Sieg, H. (2013). Estimating hedonic functions for rents and values in the presence of unobserved heterogeneity in the quality for housing. GSIA Working Papers Series No. 2013-E34. Electronic version. Pittsburgh, PA: Carnegie Mellon Tepper School of Business.

Evans, M. F. & Smith, V. K. (2006). Do we really understand the age-VSL relationship? Resource and Energy Economics, 28, 242-261.

Freeman, A. M., III, Herriges, J. A. & Kling, C. L. (2014). The measurement of environmental and resource values: Theory and Methods. Washington, DC: RFF Press.

Gamper-Rabindran, S. & Timmins, C. (2013). Does cleanup of hazardous waste sites raise housing values? Evidence of spatially localized benefits. Journal of Environmental Economics and Management, 65. 345-360.

Garber, S. & Klepper, S. (1980). Extending the classical normal errors-in-variables model. Econometrica, 48, 1541-1546.

Gawande, K., Jenkins-Smith, H. & Yuan, M. (2013). The long-run impact of nuclear waste shipments on the property market: Evidence from a quasi-experiment. Journal of Environmental Economics and Management, 65, 56-73.

Gayer, T., Hamilton, J. T. & Viscusi, W. K. (2000). Private values of risk tradeoffs at superfund sites: Housing market evidence on learning about risk. Review of Economics and Statistics, 82, 439-451.

Goodman, A. C. & Thibodeau, T. G. (2007). The spatial proximity of metropolitan area housing submarkets. Real Estate Economics, 35, 209-232.

Gopalakrishnan, S. & Klaiber, H. A. (2014). Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania. American Journal of Agricultural Economics, 96, 43-66. DOI:10.1093/ajae/aat065.

Grainger, C. A. (2012). The distributional effects of pollution regulations: Do renters fully pay for cleaner air? Journal of Public Economics, 96, 840-852.

Greene, W. H. (2011). Econometric Analysis (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Greenstone, M. & Gallagher, J. (2008). Does hazardous waste matter? Evidence from the housing market and the superfund program. The Quarterly Journal of Economics, 123, 951-1003.

Hallstrom, D. G. & Smith, V. K. (2005). Market responses to hurricanes. Journal of Environmental Economics and Management, 50, 541-561.

Halvorsen, R. & Palmquist, R. (1980). The interpretation of dummy variables in semilogarithmic equations. American Economic Review, 70, 474-475.

Hanna, B. G. (2007). House values, incomes, and industrial pollution. Journal of Environmental Economics and Management, 54, 100-112.

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46, 1251-1271.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. American Economic Review, 71, 662-676.

Heintzelman, M. D. & Tuttle, C. M. (2012). Values in the wind: A hedonic analysis of wind power facilities. Land Economics, 88, 571-588.

Hess, D. B. & Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. Urban Studies, 44, 1041-1068.

Hite, D., Chern, W., Hitzhusen, F. & Randall, A. (2001). Property-value impacts of an environmental disamenity: The case of landfills. Journal of Real Estate Finance and Economics, 22, 185-202.

Huang, J.-C. & Palmquist, R. B. (2001). Environmental conditions, reservation prices, and time on the market for housing. Journal of Real Estate Finance and Economics, 22, 203-219.

Ihlanfeldt, K. R. & Taylor, L. O. (2004). Externality effects of small-scale hazardous waste sites: Evidence from urban commercial property markets. Journal of Environmental Economics and Management, 47, 117-139.

Irwin, E. G. & Bockstael, N. E. (2001). The problem of identifying land use spillovers: Measuring the effects of open space on residential property values. American Journal of Agricultural Economics, 83, 698-704.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. American Economic Review, 72, 801.

Kiel, K. A. & McClain, K. T. (1995). Housing prices during siting decision stages: The case of an incinerator from rumor through operation. Journal of Environmental Economics and Management, 28, 241-255.

Kiel, K. A. & Williams, M. (2007). The impact of Superfund sites on local property values: Are all sites the same? Journal of Urban Economics, 61, 170-192.

Kiel, K. A. & Zabel, J. E. (1999). The accuracy of owner-provided house values: The 1978-1991 American Housing Survey. Real Estate Economics, 27, 263-398.

Kim, C. W., Phipps, T. T. & Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. Journal of Environmental Economics and Management, 45, 24-39.

Kim, J. & Goldsmith, P. (2009). A spatial hedonic approach to assess the impact of swine production on residential property values. Environmental and Resource Economics, 42, 509-534.

Klaiber, H. A. & Kuminoff, N. V. (2014). Equilibrium sorting models of land use and residential choice. In J. M. Duke & J. Wu (Eds.), The Oxford handbook of land economics (pp. 352-379). Oxford, United Kingdom: Oxford University Press.

Klaiber, H. A. & Phaneuf, D. J. (2010). Valuing open space in a residential sorting model of the Twin Cities. Journal of Environmental Economics and Management, 60, 57-77.

Klaiber, H. A. & Smith, V. K. (2013). Quasi experiments, hedonic models, and estimating trade-offs for local amenities. Land Economics, 89, 413-431.

Kniesner, T. J., Viscusi, W. K., Woock, C. & Ziliak, J. P. (2012). The value of a statistical life: Evidence from panel data. The Review of Economics and Statistics, 94, 74-87.

Kochi I., Hubbell, B. & Kramer, R. (2006). An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis. Environmental & Resource Economics, 34, 385-406.

Kochi, I. & Taylor, L. O. (2011). Risk heterogeneity and the value of reducing fatal risks: Further market-based evidence. Journal of Benefit-Cost Analysis, 2, 1-28.

Kohlhase, J. E. (1991). The impact of toxic waste sites on housing values. Journal of Urban Economics, 30, 1-26.

Kuminoff, N. V., Parmeter, C. F. & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? Journal of Environmental Economics and Management, 60, 145-160.

Kuminoff, N. V. & Pope, J. C. (2012). A novel approach to identifying hedonic demand parameters. Economics Letters, 116, 374-376.

Kuminoff, N. V. & Pope, J. C. (2013). The value of residential land and structures during the great housing boom and bust. Land Economics, 89, 1-29.

Kuminoff, N. V. & Pope, J. C. (2014). Do 'capitalization effects' for public goods reveal the public's willingness to pay? International Economic Review, 55, 1227-1250.

Kuminoff, N. V., Smith, V. K. & Timmins, C. (2013). The new economics of equilibrium sorting and policy evaluation using housing markets. Journal of Economic Literature, 51, 1007-1062.

Landry, C. E. & Hindsley, P. (2011). Valuing beach quality with hedonic property models. Land Economics, 87, 92-108.

Leamer, E. (2007). Housing IS the business cycle. Proceedings, Federal Reserve Bank of Kansas City (pp. 149-233).

Lee, J. & Taylor, L. O. (2015). Randomized safety inspections and risk exposure on the job: quasi-experimental estimates of the value of a statistical life. Working Paper. Center for Environmental and Resource Economic Policy, North Carolina State University.

Leggett, C. G. & Bockstael, N. E. (2000). Evidence of the effects of water quality on residential land prices. Journal of Environmental Economics and Management, 39, 121-144.

León, G. & Miguel, E. (2013). Transportation choices and the value of statistical life. National Bureau of Economic Research Working Paper 19494. Cambridge, MA: NBER. Last accessed Nov. 21, 2013, from http://emiguel.econ.berkeley.edu/research/transportation-choices-and-the-value-of-statistical-life-in-africa.

LeSage, J. & Pace, R. K. (2009). Introduction to spatial econometrics. Boca Raton, FL: Chapman & Hall/CRC.

Linneman, P. (1980). Some empirical results on the nature of the hedonic price function for the urban housing market. Journal of Urban Economics 8, 47-68.

Liu, X. & Lynch, L. (2011). Do agricultural land preservation programs reduce farmland loss? Evidence from a propensity score matching estimator. Land Economics, 87, 183-201.

Liu, X., Taylor, L. O., Hamilton, T. L. & Grigelis, P. E. (2013). Amenity values of proximity to National Wildlife Refuges: An analysis of urban residential property values. Ecological Economics, 94, 37-43.

Matthews, J. W. & Turnbull, G. K. (2007). Neighborhood street layout and property value: The interaction of accessibility and land use mix. Journal of Real Estate Finance and Economics, 35, 111-141.

McConnell, V. & Walls, M. (2005). The value of open space: Evidence from studies of nonmarket benefits. Washington, DC: Resources for the Future.

Meyer, B. D. (1995). Natural and quasi-experiments in economics. Journal of Business and Economic Statistics, 13, 151-61.

Moffit, R. (1991). Program evaluation with nonexperimental data. Evaluation Review, 15, 291-314.

Mrozek, J. R. & Taylor, L. O. (2002). What determines the value of life? A meta-analysis. Journal of Policy Analysis and Management, 21, 253-270.

Nakamura, A. & Nakamura, M. (1998). Model specification and endogeneity. Journal of Econometrics, 83, 213-237.

Noonan, D. S., Krupka, D. J. & Baden, B. M. (2007). Neighborhood dynamics and price effects of superfund site clean-up. Journal of Regional Science, 47, 665-692.

Ohta, M. & Griliches, Z. (1976). Automobile prices revisited: Extensions of the hedonic hypothesis. In N. Terleckyj (Ed.), Household production and consumption (pp. 325-398). Cambridge, MA: National Bureau of Economic Research.

Palmquist, R. B. (1984). Estimating the demand for the characteristics of housing. The Review of Economics and Statistics, 66, 394-404.

Palmquist, R. B. (1988). Welfare measurement for environmental improvements using the hedonic model: The case of nonparametric marginal prices. Journal of Environmental Economics and Management, 15, 297-312.

Palmquist, R. B. (1991). Hedonic methods. In J. B. Braden & C. D. Kolstad (Eds.), Measuring the demand for environmental quality (pp. 77-120). North-Holland, Amsterdam, The Netherlands: North-Holland.

Palmquist, R. B. (1992a). Valuing localized externalities. Journal of Urban Economics, 31, 59-68.

Palmquist, R. B. (1992b). A note on transactions costs, moving costs, and benefit measurement. Journal of Urban Economics, 32, 40-44.

Palmquist, R. B. (2006). Property value models. In K.-G. Mäler & J. R. Vincent (Eds.), Handbook of environmental economics, Volume 2: Valuing environmental changes (pp. 763-819). Amsterdam, The Netherlands: Elsevier.

Palmquist, R. B. & Israngkura, A. (1999). Valuing air quality with hedonic and discrete choice models. American Journal of Agricultural Economics, 81, 1128-1133.

Parmeter, C. F. & Pope, J. C. (2013). Quasi-experiments and hedonic property value methods. In J. A. List & M. K. Price (Eds.), Handbook of experimental economics and the environment (pp. 3-65). Cheltenham, United Kingdom: Edward Elgar.

Parsons, G. R. (1986). An almost ideal demand system for housing attributes. Southern Economic Journal, 53, 347-363.

Paterson, R. W. & Boyle, K. J. (2002). Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. Land Economics, 78, 417-425.

Petrie, R. A. & Taylor, L. O. (2007). Estimating the value of water use permits: A hedonic approach applied to farmland in the southeastern United States. Land Economics, 83, 302-318.

Phaneuf, D.J. & Requate, T. (in press). A course in enviornmental economics: Theory, policy, and practice. New York: Cambridge University Press.

Phaneuf, D. J., Taylor, L. O. & Braden, J. B. (2013). Combining revealed and stated preference data to estimate preferences for residential amenities: A GMM approach. Land Economics, 89, 30-52.

Poor, P. J., Boyle, K. J., Taylor, L. O. & Bouchard, R. (2001). Objective versus subjective measures of water quality in hedonic property value models. Land Economics, 77, 482-493.

Pope, J. C. (2008). Buyer information and the hedonic: The impact of a seller disclosure on the implicit price for airport noise. Journal of Urban Economics, 63, 498-516.

Quigley, J. M. (1982). Nonlinear budget constraints and consumer demand: An application to public programs for residential housing. Journal of Urban Economics, 12, 177-201.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy 82, 34-55.

Sander, H., Polasky, S. & Haight, R. G. (2010). The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA. Ecological Economics, 69, 1646-1656.

Sanders, N. J. (2012). Toxic assets: How the housing market responds to environmental information shocks. Working Paper 128. Williamsburg, VA: College of William and Mary, Department of Economics. Retrieved Nov. 2, 2103, from IDEAS database.

Scotton, C. R. & Taylor, L. O. (2011). Valuing risk reductions: Incorporating risk heterogeneity into a revealed preference framework. Resource and Energy Economics, 33, 381-397.

Shiller, R. J. (2007). Understanding recent trends in house prices and home ownership. National Bureau of Economic Research Working Paper 13553. Cambridge, MA: NBER.

Sieg, H., Smith, V. K., Banzhaf, H. S. & Walsh, R. (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. International Economic Review, 45, 1047-1077.

Small, K. A. & Steimetz, S. S. C. (2012). Spatial hedonics and the willingness to pay for residential amenities. Journal of Regional Science, 52, 635-647.

Smith, V. K., Evans, M. F., Kim, H. & Taylor, D. H. Jr. (2004). Do the near-elderly value mortality risks differently? Review of Economics and Statistics, 86, 423-29.

Smith, V. K. & Huang, J.-C. (1995). Can markets value air quality? A meta-analysis of hedonic property value models. Journal of Political Economy, 103, 209-227.

Taylor, L. O. (2008). Theoretical foundations and empirical developments in hedonic modeling. In A. Baranzini, J. Ramirez, C. Schaerer & P. Thalmann (Eds.), Hedonic methods in housing markets: Pricing environmental amenities and segregation (pp. 15-37). New York: Springer.

Taylor, L. O. Phaneuf, D. J. & Liu, X. (2016). Disentangling the property value impacts of environmental contamination from locally undesirable land uses: Implications for measuring post-cleanup stigma, Journal of Urban Economics, 93, 85-98.

Taylor, L. O. & Smith, V. K. (2000). Environmental amenities as a source of market power. Land Economics, 76, 550-568.

Todd, P. E. (2008). Matching estimators. From: The New Palgrave Dictionary of Economics. In S. N. Durlauf & L. E. Blume (Eds.). The New Palgrave Dictionary of Economics Online (2nd ed.). Palgrave Macmillan. Retrieved from www.dictionaryofeconomics.com/article?id=pde2008_M000365.

Tra, C. I. (2010). A discrete choice equilibrium approach to valuing large environmental changes. Journal of Public Economics, 94, 183-196.

Tra, C. I. (2013). Measuring the general equilibrium benefits of air quality regulation in small urban areas. Land Economics, 89, 291-307.

U.S. Environmental Protection Agency. (1997). The benefits and costs of the Clean Air Act: 1970 to 1990. Prepared by the Office of Administration and Resources Management, Office of Policy, Planning, and Evaluation for the U.S. Congress, October 1997.

U.S. Environmental Protection Agency. (1999). The benefits and costs of the Clean Air Act: 1990 to 2010. Prepared by the Office of Air and Radiation, Office of Policy for the U.S. Congress, November 1999.

Vartia, Y. O. (1983). Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions. Econometrica, 51, 79-98.

Viscusi, W. K. (1992). Fatal tradeoffs: Public and private responsibilities for risk. New York: Oxford University Press.

Viscusi, W. K. (1993). The value of risks to life and health. Journal of Economic Literature, 31, 1912-1946.

Viscusi, W. K. & Aldy, J. E. (2003). The value of a statistical life: A critical review of market estimates throughout the world. Journal of Risk and Uncertainty, 27, 5-76.

Waugh, F. V. (1928). Quality factors influencing vegetable prices. Journal of Farm Economics, 10, 185-196.

Zabel, J. (2013). Hedonic models and the housing cycle. Working Paper, Department of Economics. Medford, MA: Tufts University.

Zabel, J. E. & Guignet, D. (2012). A hedonic analysis of the impact of LUST sites on house prices. Resource and Energy Economics, 34, 549-564.

Zabel, J. E., and Kiel, K. A. (2000). Estimating the demand for air quality in four U.S. cities. Land Economics, 76, 174-194.

Chapter 8
Averting Behavior Methods

Mark Dickie

Abstract Averting behavior refers to actions taken to defend against environmental or other hazards, whether by reducing exposure to hazards or by mitigating adverse effects of exposure. This chapter examines how information on averting behavior can be used to improve nonmarket valuation. The chapter describes the history, theoretical foundation, and empirical application of averting behavior methods. These methods support estimation of economic benefits of public policies, especially those that reduce morbidity or mortality, in a way that is consistent with utility maximization and is based on observable behavior. The chapter: (1) shows how ignoring averting behavior may cause an invalid measurement of physical and economic damages of pollution or other hazards, and how controlling for averting behavior may improve welfare measurement; (2) explains several ways of using information on averting behavior to estimate the benefits of environmental improvement; (3) provides a simple empirical illustration; and (4) argues that the validity of welfare measurement using averting behavior methods depends on how the challenges of joint production, unknown prices of averting actions, and identification of the effects of pollution and averting behavior are met.

Keywords Averting · Avoidance · Defensive · Mitigating · Behavior · Expenditure · Self-protection · Value of statistical life · Mortality · Morbidity · Health · Valuation · Willingness-to-pay · Damage function

Individuals facing environmental or other hazards sometimes act to reduce the damage they may suffer. They may reduce exposure to ambient air pollution by adjusting daily activities or reduce exposure to contaminated water supplies by changing sources of drinking water. Aside from avoiding exposure, people may act to mitigate the damage caused by a given amount of exposure. They may use medical treatment or medication to offset adverse health effects of pollution exposure or use seat belts, child safety seats, motorcycle helmets, or bicycle helmets

M. Dickie (✉)
University of Central Florida, Orlando, USA
e-mail: mdickie@ucf.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_8

to reduce the risk of death from traffic hazards. "Averting behavior" (or avoidance, defensive, mitigating, or protective behavior) refers to actions taken to defend against hazards, whether by reducing exposure or by offsetting the effects of exposure.

A rational person facing pollution or other hazards will act to defend himself as long as he perceives the benefits of defensive action to be greater than the costs. This simple observation has two main implications for nonmarket valuation. First, ignoring averting behavior may cause invalid measurement of the physical and economic damages of pollution. Second, accounting for averting behavior may provide information about the benefits of environmental improvement.

To illustrate the two implications of protective behavior for nonmarket valuation, imagine a temporarily contaminated supply of drinking water. Suppose that in the absence of avoidance behavior, the contamination would increase the number of cases of diarrheal illness by 4% among exposed persons. Suppose that half of the affected population does nothing to avoid the contamination and consequently experiences the 4% increase in illness. The other half of the population avoids the contamination by switching to another source of drinking water and experiences no increase in illness. The number of cases of illness in the total population increases by 2%. If averting behavior is ignored, the physical damage of contamination is understated because the apparent increase in cases of illness (2%) understates the true increase that would have been caused if behavior were unchanged (4%). The economic damage is also understated because the costs of avoiding contaminated water constitute part of the economic damage, but avoidance costs are not counted if averting behavior is ignored. Ignoring averting behavior when it is present causes an invalid measurement of the physical and economic damages of contamination.

Suppose further that avoiding the contaminated water supply requires $20 in money and time costs. The contamination increased the probability of illness by 4%, and the averting expenditure of $20 was just sufficient to fully offset the increased risk of illness for those who averted. If everyone faced the same $20 cost of avoiding the contamination, then the half of the population that averted revealed that they are willing to pay at least $20 to avoid a 4% increase in the risk of diarrheal illness. The half that did not avert revealed that they are willing to pay less than $20. Median willingness to pay is $20. Thus, accounting for the effect of contamination on health, the effect of averting behavior on health and the cost of averting action provides information about the economic benefits of avoiding contaminated water.

Economists have applied several variations to the logic just described to investigate the benefits of pollution control. They have estimated benefits of avoiding water contamination based on costs of averting actions such as purchasing bottled water, boiling or purifying water, hauling water from an alternate source, or installing water filtration systems (Harrington et al. 1989; Abdalla 1990; Dasgupta 2004). They have estimated benefits of reduced air pollution by examining medical care that mitigates adverse health effects of air pollution exposure (Joyce et al. 1989; Gerking and Stanley 1986; Dickie and Gerking 1991b); use of home air cleaners, purifiers, or conditioners (Dickie and Gerking 1991a; Richardson et al. 2012); or changes in daily activities (Neidell 2009). They have estimated the benefits of reducing risks of skin cancer based on averting expenditures for sunscreen lotion (Murdoch and Thayer 1990) and have examined expenditures on energy consumption for residential cooling to offset increases in heat-related mortality induced by climate change (Deschenes and Greenstone 2011). In related work, economists have estimated the benefits of reduced risk of death by considering consumers' purchase and use of safety equipment like motorcycle and bicycle helmets (Blomquist 2004). As these examples illustrate, the majority of applications have involved human health, but other applications include household cleaning activities to reduce soiling damages from air pollution (Watson and Jaksch 1982) and actions taken to defend against pest infestation (Jakus 1994).

Although averting behavior methods have been applied widely, three challenges hinder valid measurement of benefits. First, joint production occurs when averting behavior jointly produces additional benefits or costs beyond its impact on mitigating pollution damages. Valuation requires isolating the benefits of mitigation from other benefits and costs, but doing so is difficult in the presence of joint production. Second, measuring benefits requires knowing the cost of averting behavior, but the prices of many defensive actions are not easily measured. Third, valuation requires identifying the effects of pollution and/or of averting behavior on health or on other outcomes that people value, but identification of either effect is challenging.

This chapter examines how information on averting behavior can be used to improve nonmarket valuation. The chapter has three main objectives: (1) to show how ignoring averting behavior may cause an invalid measurement of physical and economic damages of pollution, and how controlling for averting behavior may improve welfare measurement; (2) to explain several ways of using information on averting behavior to estimate the benefits of environmental improvement; and (3) to argue that the validity of welfare measurement using averting behavior methods depends on how the challenges of joint production, unknown prices of averting actions, and identification of the effects of pollution and averting behavior are met.

8.1 Historical Development of Averting Behavior Methods

Averting behavior methods were developed to support estimation of the economic benefits of public policies, particularly policies that reduce premature mortality or morbidity, in a way that is consistent with the theory of utility maximization and is based on observable behavior.[1] Benefits of pollution control are typically estimated

---

[1]Although most research on averting behavior focuses on nonmarket valuation, economists also have investigated consequences of averting behavior for efficient control of externalities. Three key issues considered are (1) whether or not persons damaged by an externality who have opportunities for averting action should be taxed or compensated to achieve efficiency (Coase 1960; Baumol and Oates 1988); (2) how averting behavior affects the efficient quantity of an

using the two-step damage function method (Freeman 1979; U.S. EPA 2012, Section 5.4). The first step of the damage function method is to estimate a physical damage function or dose-response function that specifies physical damage—such as crop loss, depreciation or soiling of materials, or loss of health or life—as a function of pollution concentrations and other variables. The damage function is used to predict the reduction in physical damage associated with a given reduction in pollution. The second step of the damage function method is to determine an economic value per unit of physical damage. Benefits of pollution control are computed as the product of the predicted reduction in physical damage and the unit value.

The damage function method suffers from at least two potential shortcomings. First, it treats economic agents as passive receptors of damage rather than as rational actors who would defend against pollution when the benefits of averting action exceed the costs. Thus, damage functions as typically applied are inconsistent with maximizing behavior by persons who have opportunities for averting behavior. In fact, for many environmental insults there are averting actions that would reduce damages (Courant and Porter 1981). Farmers suffering crop damage from pollution might reduce the damage through costly input or output substitutions (Garcia et al. 1986); property owners facing damage to materials might increase maintenance or cleaning expenditures (Watson and Jaksch 1982); individuals exposed to pollution might engage in preventive activities that mitigate adverse health effects of exposure (Cropper 1981). By failing to account for the effects of averting behavior on physical damages, the damage function method may underestimate the amount of physical damage that would result from unmitigated increases in pollution. And by failing to account for the costs of preventive measures, the damage function method may underestimate the unit value of reduced damage.

A second major shortcoming of the damage function method, particularly as applied when averting behavior methods were initially developed, is that the economic value per unit of physical damage that is applied is often not consistent with the principles of welfare economics. In the case of reduced morbidity or premature mortality, benefits historically were measured using the human capital or cost of illness approach based on earnings forgone from illness or premature death plus treatment costs for illness (Mushkin and Collings 1959; Weisbrod 1971; Rice 1966; Dickie 2003). A measure of benefit that is consistent with the Kaldor–Hicks criteria of welfare economics, in contrast, would be based on individual willingness to pay for reduced morbidity or for reduced risk of mortality or morbidity (Schelling 1968; Mishan 1971). Forgone earnings and medical expenses do not equal willingness to pay. The human capital and cost of illness measures have little basis in the theory of

(Footnote 1 continued)

externality (Zeckhauser and Fisher 1976; Oates 1983) and/or complicates policy design and implementation by causing nonconvexity of the social marginal damage function (Shibata and Winrich 1983; McKitrick and Collinge 2002); and (3) whether providing information about environmental, health, or safety risks can reduce damages by promoting averting action (Smith et al. 1995; Shimshack et al. 2007; Graff Zivin and Neidell 2009).

utility-maximizing behavior and when used as welfare measures are inconsistent with the principles of welfare economics.

As the shortcomings of existing valuation procedures became apparent, policy developments in the United States heightened the importance of developing better methods for estimating benefits of improved health and reduced pollution. During the 1970s, the U.S. federal government substantially expanded its regulatory efforts to reduce health and safety hazards in the environment and workplaces and for consumer products. Concern for the economic effects of increased regulation prompted requirements for assessments of economic impacts. For instance, Executive Order 12291, issued in 1981, required assessment of the benefits and costs of all major proposed federal regulations (Harrington and Morgenstern 2004). With much of the growth in regulation focused on reducing health and safety risks, the need for improved measurement of the benefits of reduced morbidity and mortality was pressing.

Averting behavior methods emerged from this setting.[2] The methods were designed to support the estimation of willingness to pay based on observable behavior and/or to improve the measurement of the physical effects of pollution by accounting for behavioral responses to changes in environmental quality.

Four key contributions guided development of methods. First, three papers used averting behavior models to estimate welfare changes. Blomquist (1979) established that willingness to pay for reduced mortality risk could be inferred from observed averting behavior. Cropper (1981) demonstrated that accounting for preventive averting behavior could support alternatives to the damage function approach for estimating benefits of reduced morbidity from pollution control. Watson and Jaksch (1982) illustrated estimation of willingness to pay for reduced air pollution based on averting expenditures. Second, the theoretical basis for estimating willingness to pay based on averting behavior and the relationship between willingness to pay and other measures of value were developed systematically (Courant and Porter 1981; Bockstael and McConnell 1983; Harrington and Portney 1987; Bartik 1988). Third, economists recognized identification challenges in the averting behavior model and developed methods to meet these challenges (Gerking and Stanley 1986; Harrington and Portney 1987; Neidell 2009; Deschenes and Greenstone 2011). Fourth, researchers accumulated empirical evidence that a substantial proportion of people take averting actions, that averting behavior increases with increases in ambient pollution or other hazards, and that ignoring averting behavior causes an understatement of the physical and economic damages of pollution.[3]

---

[2]Against the same background, economists interested in improving methods for estimating benefits of reduced mortality also applied the contingent valuation method (Jones-Lee et al. 1985) and developed methods for inferring the value of reduced mortality from estimated trade-offs between wages and risks (Thaler and Rosen 1976).

[3]Early research provided evidence that (1) large proportions of survey respondents report that they change their behavior when pollution is high (Berger et al. 1987; Bresnahan et al. 1997); (2) many people take a variety of defensive actions in response to water contamination (Harrington et al.

At first glance, averting behavior methods might seem ideally suited for application to important nonmarket valuation problems. The method emphasizes valuation of reduced mortality and morbidity, and estimated benefits of reduced mortality and morbidity often account for a large majority of the measured economic benefits of pollution control.[4] On the contrary, however, averting behavior methods have made few inroads into the world of applied policy analysis.

The damage function approach remains the standard method for assessing the benefits of environmental programs (U.S. EPA 2012). Mortality and morbidity impacts of pollution are measured using damage functions that fail to control for averting behavior. The unit values of mortality and morbidity reductions are derived from wage-risk trade-offs, stated-preference studies, or costs of illness, not from studies of averting behavior. For example, the U.S. relies mainly on estimated wage-risk trade-offs to measure the benefits of reduced mortality, the European Commission and Australia rely mainly on stated-preference values, and the United Kingdom makes use of both types of valuation estimates (OECD 2011). Benefits of reduced morbidity are based on stated preference or, more frequently, on cost of illness values. In short, the two insights of the averting behavior method—that accounting for averting behavior can improve measurement of the physical effects of pollution and provide information on the economic benefits of pollution control—have had little impact on policy analysis. One reason for the limited impact of averting behavior methods on policy analysis is concern that challenges confronting the method, particularly joint production, unknown prices for averting actions, and identification of the effects of pollution and of averting behavior are not addressed satisfactorily in many studies.

(Footnote 3 continued)

1989; Abdalla 1990); (3) daily activities and demand for medical care respond to changes in ambient air pollution (Dickie and Gerking 1991b; Bresnahan et al. 1997); (4) mitigation increases with residential radon concentrations (Akerman et al. 1991; Doyle et al. 1991; Smith et al. 1995); and (5) individuals whose personal characteristics make them more susceptible to environmental hazards are more likely than others to engage in protective behavior (Bresnahan et al. 1997; Dickie and Gerking 1997, 2009). More recent work has examined behavioral responses to public information about air pollution hazards and/or has used quasi-experimental methods to identify causal effects of pollution, or information about pollution, on behavior and health. These studies support the inference that people act to reduce exposure to pollution hazards (Neidell 2009), implying that ignoring averting behavior causes understatement of the health effects of pollution and of the benefits of pollution reduction.

[4] For example, in the retrospective analysis of benefits and costs of the U.S. Clean Air Act, benefits of reduced mortality accounted for more than 80% of the total estimated benefits, and benefits of reduced morbidity made up most of the remainder (U.S. EPA 1997).

Table 8.1 Data from a hypothetical natural experiment

	Day 1			Day 2					
	Low air pollution			**Low air pollution**			**High air pollution**		
Number of children who:	Healthy	Sick	Total	Healthy	Sick	Total	Healthy	Sick	Total
A. Children who are less sensitive to health effects of pollution									
Play outdoors	200	0	200	100	0	100	64	16	80
Play indoors	0	0	0	0	0	0	20	0	20
Total	200	0	200	100	0	100	84	16	100
B. Children who are more sensitive to health effects of pollution									
Play outdoors	200	0	200	100	0	100	12	8	20
Play indoors	0	0	0	0	0	0	64	16	80
Total	200	0	200	100	0	100	76	24	100
C. All children									
Play outdoors	400	0	400	200	0	200	76	24	100
Play indoors	0	0	0	0	0	0	84	16	100
Total	400	0	400	200	0	200	160	40	200

8.2 How Averting Behavior Affects Nonmarket Valuation

This section illustrates how averting behavior affects nonmarket valuation, without resorting to formal economic theory or econometrics using three examples based on a hypothetical natural experiment. The first example shows how controlling for averting behavior can improve measurement of the physical and economic damages of pollution. The second example shows how accounting for averting action provides information on the economic benefits of environmental improvements and shows how joint production or unobserved prices of averting behavior threaten valid measurement of benefits. The third example illustrates how the challenges of identifying the causal effects of pollution and of averting behavior can threaten the validity of inferences about the physical and economic damages of pollution.

Consider the following hypothetical natural experiment. Each child in a population of 200 children has the same resistance to illness and the same propensity to experience illness from exposure to a given concentration of an ambient air pollutant, such as ozone. On one summer day when ambient ozone levels are low, each child plays outdoors and none is sick. On a second summer day, ambient ozone levels remain low in the area where one-half of the children are located, but ambient levels are high where the other half of the children are located. Assume that the increase in ambient ozone is "assigned as if randomly"; in other words, the ozone concentration is independent of all observed and unobserved characteristics of the children.

Suppose that all of the 100 children facing low pollution on the second day continue to play outdoors, and none is sick. Among the 100 children facing high pollution on the second day, 80 play outdoors and 16 of them experience ozone-induced illness. The remaining 20 children stay indoors—perhaps their parents kept them indoors in response to an air quality alert—and none is sick. (Because ozone breaks down rapidly indoors, limiting outdoor activities is an effective way to avoid ozone pollution.) These data are summarized in Panel A of Table 8.1.

8.2.1 Controlling for Averting Behavior to Measure Effects of Pollution

To measure the economic benefits of avoiding increased ambient ozone in the population of 200 children described in Panel A of Table 8.1, consider applying the two-step damage function method. First, measure the increase in illness associated with the increase in pollution while ignoring averting behavior (that is, while making no distinction between children who play indoors or outdoors). Second, multiply the increase in illness computed in Step 1 by a unit value of avoiding illness.

To implement Step 1, assume that the effectively random assignment of ambient air pollution creates a natural or quasi-experiment in which the "control group" consists of children who do not experience the increase in ambient air pollution. The "treatment group" consists of children who experience the increase in pollution on Day 2. Compare the difference in illness between Day 2 and Day 1 that occurred in the treatment group to the difference that occurred in the control group.

Observing that 16 of the 100 children facing high pollution on Day 2 became sick, compared to none who were sick on Day 1, the difference in the probability of illness is $16/100 - 0/100 = 0.16$ in the treatment group. In contrast, there was no increase in illness between Day 1 and Day 2 among the 100 children who faced low pollution on both days. The difference in probability of illness is $0/100 - 0/100 = 0$ in the control group. A difference-in-differences measure of the effect of elevated ozone on the probability of illness is $(0.16 - 0) - (0 - 0) = 0.16$. If all 200 children were expected to experience higher ozone concentrations, Step 1 of the damage function method predicts $0.16 \times 200 = 32$ additional cases of illness.

In Step 2, multiply the predicted number of additional cases of illness by an economic value per avoided illness. For instance, if previous studies indicate that the value of avoiding a day of childhood illness is \$100, the estimated economic benefit of avoiding the ozone increase is $\$100 \times 32 = \$3,200$ for the population of children described in Table 8.1, Panel A.

Unfortunately, by failing to control for averting behavior (staying indoors), the damage function method understates the effect of pollution on illness, the expected number of additional cases of illness, and the economic benefit of pollution control. Note that not all of the children living in the high ozone area actually experienced the increase in pollution because some of them were kept indoors. To measure the causal effect of pollution, everything except pollution should be held constant; "everything" includes the behavior of the people exposed to pollution. But the behavioral response to increased pollution was not controlled in Step 1 of the damage function method. To hold behavior constant despite the increase in pollution, consider a treatment group that consists only of children who play outdoors on both days.[5]

Observing that 16 of 80 children who faced increased pollution on Day 2 and played outdoors on both days became sick, compared to none who were sick on Day 1, the difference in probability of illness among this treatment group is $16/80 - 0/80 = 0.2$. Thus, the difference-in-differences measure of the causal effect of pollution, controlling for behavior, is $(0.2 - 0) - (0 - 0) = 0.2$. Assuming again that all 200 children would face a comparable increase in ozone, an additional $0.2 \times 200 = 40$ cases of illness would be expected *if behavior were held constant*. Applying the same value of \$100 per day of illness, the estimated benefits of avoiding the pollution increase are $\$100 \times 40 = \$4,000$ for the population of children described in Panel A of Table 8.1.

In this example, failing to control for averting behavior caused an invalid measurement of the effect of pollution on the probability of illness in the amount of $0.16 - 0.2 = -0.04$, an understatement of 20% of the true causal effect. The expected number of additional cases of illness and the economic benefit of pollution control were understated by 20% as well.

Notice that the invalid measurement that occurs when averting behavior is ignored does not result from making statistical inferences about a population based on observations of a sample. Instead, the invalid measurement results from measuring the wrong quantity. The damage function measures the total effect of pollution rather than the partial effect. The distinction between the partial and total effects of pollution is central to the averting behavior method and is worth examining in detail.

The partial effect measures the impact of pollution with behavior held constant. Given a random assignment of pollution and constant behavior, the partial effect measures the causal effect of pollution. In contrast, the total effect measures the combined effect of pollution and any behavioral changes induced by a change in pollution. The difference between the total and partial effects of pollution equals the product of (1) the effect of pollution on averting behavior and (2) the effect of averting behavior on health. Consider measuring these two effects using data in Panel A of Table 8.1.

To measure the effect of pollution on defensive behavior, note that 20 of the 100 children who faced an increase in pollution changed their behavior: 20/100 stayed indoors when pollution was high, compared to 0/100 when pollution was low. The difference in probability of avoidance behavior among those facing increased

[5]Behavior is already held constant in the control group of children who experience no increase in pollution because these children play outdoors both days. The difference in probability of illness in this group is the same as measured originally: $0/100 - 0/100 = 0$.

pollution is 0.2. In contrast, none of the children who faced no increase in pollution changed their behavior. The difference in probability of avoidance behavior among this group is 0. The difference-in-differences measure of the effect of the randomly assigned increase in pollution on the probability of averting behavior is $0.2 - 0 = 0.2$.

To measure the effect of averting behavior on health, one should focus on the change in the probability of becoming sick when behavior changes, while holding pollution constant. Consider the children who faced higher pollution and stayed indoors as the group choosing the "treatment" of averting behavior, and the children who faced higher pollution but played outdoors as the control group that did not choose the treatment.[6] There is no difference in illness between Day 1 and Day 2 among the 20 children who faced high ozone on Day 2 but remained indoors (Table 8.1, Panel A). The change in probability of illness in the presence of averting is $0/20 - 0/20 = 0$. Among the 80 children who faced high ozone and played outdoors, 16 children became sick, compared to none who were sick when facing low ozone levels. The difference in the probability of illness in the absence of averting is $16/80 - 0/80 = 0.2$. Relying on the crucial assumption that all of the children are equally resistant to illness, the difference-in-differences measure of the effect of averting behavior on the probability of illness, with pollution held constant, is $0 - 0.2 = -0.2$. In other words, staying indoors reduces the probability of illness by 0.2 when ozone levels are high.

Equation (8.1) summarizes the relationship between total and partial effects of pollution.

$$\begin{array}{ccccccc} \begin{pmatrix} \text{Total effect} \\ \text{of pollution} \\ \text{on illness} \end{pmatrix} & = & \begin{pmatrix} \text{Partial effect} \\ \text{of pollution} \\ \text{on illness} \end{pmatrix} & + & \begin{pmatrix} \text{Effect of} \\ \text{averting} \\ \text{on illness} \end{pmatrix} & \times & \begin{pmatrix} \text{Effect of} \\ \text{pollution on} \\ \text{averting} \end{pmatrix} \\ 0.16 & = & 0.2 & + & (-0.2) & \times & 0.2. \end{array} \tag{8.1}$$

As shown in Eq. (8.1), the total effect of pollution equals a sum of direct and indirect effects. The direct effect is the partial effect of pollution on health, holding behavior constant (0.2). The indirect effect is the effect of pollution on health that operates through the behavioral response to pollution $(-0.2 \times 0.2 = -0.04)$. The indirect effect indicates the error caused by measuring the total effect of pollution instead of the partial effect.

Ignoring averting behavior when it is present amounts to measuring the total effect of pollution rather than the partial or causal effect. As shown in Eq. (8.1), if averting behavior increases with pollution and reduces illness, then ignoring

[6]The "treatment" of averting behavior is self-selected. The difference-in-differences measure of the effect of averting behavior is valid because the children in the treated (indoors) and untreated (outdoors) groups are assumed to be identical but for the choice to avert. The difficulties arising in the more realistic situation in which the children are not identical and the differences between children are related both to health and to averting behavior are taken up in Sect. 8.2.3.

averting behavior causes the effect of pollution on illness to be understated. Equation (8.1) demonstrates that the size of the error will be small if averting behavior is not very responsive to pollution or has little impact on health. But if behavior is quite responsive to pollution and has a substantial impact on health, the error will be large. The effect of pollution on averting behavior and the effect of behavior on health are likely to vary depending on the pollutant and health effects considered and the population affected.

Neidell (2009) estimated that smog alerts reduced children's daily attendance at the Los Angeles Zoo by more than 20%, indicating a sizeable response of averting behavior to information about pollution. He estimated that the partial effect of a small increase in ambient ozone concentrations on children's hospitalizations for asthma is an increase of 2.88%, whereas the total effect is an increase of 1.09%. In this case, ignoring behavioral responses to pollution would cause a large proportionate error, because the estimated partial effect is 2.6 times larger than the total effect.

One last point is worth making before concluding this example. The increase in pollution was assumed to be randomly assigned. In reality, pollution concentrations vary spatially, and people are not randomly sorted across space. Areas with different pollution concentrations may differ in any number of other ways that affect health, making measurement of the effect of pollution on health quite challenging even in the absence of averting behavior. In recent years economists have made increasing use of quasi-experimental methods to identify the causal effect of pollution based on situations in which variation in pollution is assigned "as if randomly" (Pope 1989; Chay and Greenstone 2003; Lleras-Muney 2010). But many quasi-experimental studies do not control for averting behavior and thus do not measure partial effects of pollution.

8.2.2 Using Information on Averting Behavior to Measure Economic Benefits

To illustrate the second way that information on averting behavior can improve nonmarket valuation, consider estimating economic benefits of pollution control based on averting behavior. Suppose that keeping a child indoors imposes time and money costs on a parent or caregiver of \$20 per day. In the example just discussed, keeping a child indoors on a day when ozone concentrations are high reduces the probability of illness by 0.2. A parent who keeps her child indoors during high ozone concentrations thus reveals that she is willing to pay \$20 to reduce the probability of illness by 0.2. For every five parents who keep a child indoors, the expected number of days of child illness falls by one (5×0.2). Collectively, five parents who keep a child indoors would bear time and money costs of \$100 ($5 \times \$20$). This reasoning suggests that among parents who take averting action, willingness to pay to avoid 1 day of child illness is \$100. In this way, knowing the effect of averting behavior on illness and the cost of averting behavior allows one to estimate the value of avoiding 1 day of illness using Eq. (8.2):

$$\begin{aligned} \begin{pmatrix} \text{WTP for reduced} \\ \text{illness} \end{pmatrix} &= -\left(\frac{\text{Unit cost of averting}}{\text{Effect of averting on illness}} \right) \\ \$100 &= -\frac{\$20}{-0.2}. \end{aligned} \tag{8.2}$$

This reasoning can be extended to value pollution control using Eq. (8.3), which shows that an averting parent's willingness to pay to avoid the increase in ozone concentrations equals \$20:

$$\begin{aligned} \begin{pmatrix} \text{WTP for reduced} \\ \text{pollution} \end{pmatrix} &= \begin{pmatrix} \text{WTP for reduced} \\ \text{illness} \end{pmatrix} \times \begin{pmatrix} \text{Partial effect of} \\ \text{pollution on illness} \end{pmatrix} \\ &= -(\text{Unit cost of averting}) \times \left(\frac{\text{Partial effect of pollution on illness}}{\text{Effect of averting on illness}} \right) \\ \$20 &= -\$20 \left(\frac{0.2}{-0.2} \right). \end{aligned} \tag{8.3}$$

Equations (8.2) and (8.3) show how information on averting behavior can be used to estimate benefits, but two problems complicate application. First, staying indoors is not a market good purchased at an easily observed price. It may be difficult to estimate a "price," such as \$20, for keeping a child indoors for a day. Second, keeping a child indoors may jointly produce benefits or costs other than reducing the probability of illness. Being indoors rather than outdoors may be more or less enjoyable for the child and for the caregiver and may increase the risk of obesity from inactivity or decrease the risk of skin cancer from exposure to sunlight. The benefit measures computed in Eqs. (8.2) and (8.3) rest on the assumptions that all costs of defensive behavior were captured in the price, and all benefits were included in the reduced risk of illness. If there are additional benefits and costs, then Eqs. (8.2) and (8.3) are not valid measures of benefits. These two problems—unmeasured prices of averting actions and joint production—are common.

8.2.3 Difficulty of Identifying Partial Effects of Pollution and of Behavior

The previous two examples illustrate how properly accounting for averting behavior can improve measurement of the physical effects of pollution and can support estimation of benefits. However, valid measurement of partial effects and benefits is threatened in the likely event that an unobserved factor is related both to health and to averting behavior. To illustrate, consider a second group of children.

Each of the children in this group is equally resistant to illness, but each is more likely to experience illness than the children considered in the two previous examples. Half of the 200 children in this second group face the same randomly assigned increase in ambient ozone concentrations that was considered in Panel A of Table 8.1. Owing to their lower resistance to illness, more of the children in the second group become sick when ozone increases. In an effort to avoid this outcome, more of them are kept indoors when ozone rises. Data in Panel B of Table 8.1 can be used to determine that in this group, the total effect of increased ozone on probability of illness is 0.24, the partial effect among children who play outdoors is 0.4, increased ozone boosts the probability of averting by 0.8, and the effect of averting on the probability of illness is -0.2.[7] Using Eq. (8.3), individual willingness to pay to avoid the increased pollution in this group is $\$20(0.4/0.2) = \$40$. Table 8.2 summarizes the various effects in each group of children (from Panels A and B of Table 8.1) and the averages of the effects in the total population consisting of both groups.

Now consider estimating the partial effect of pollution on illness and the benefit of pollution control if resistance to illness is unobserved. One would be unable to stratify the population as in Panels A and B of Table 8.1 and would observe only the combined population shown in Panel C.

If differential resistance to illness is not observed, controlling for averting behavior does not necessarily support valid measurement of the partial effect of pollution on health. Using data in Panel C of Table 8.1 to attempt to measure the partial effect of pollution on the probability of illness, observe that 24 of 100 children facing high ozone levels on Day 2 and playing outdoors are sick, compared to none who were sick on Day 1. There was no increase in illness between Day 1 and Day 2 among the 200 children who faced low ambient ozone and played outdoors on both days. Given random assignment of the increase in ozone concentrations, the elevated ozone concentrations appear to cause the probability of illness to increase by $0.24 - 0 = 0.24$. However, this measure does not equal the average partial effect in the total population (0.3), or the partial effect in either group (0.2 in Group A or 0.4 in Group B). Thus, if an unobserved factor (such as

---

[7]In Panel B of Table 8.1, none of the 100 children facing low pollution on both days is sick on either day. In contrast, of the 100 children facing increased pollution on Day 2, 24 are sick when pollution is high, compared to none when pollution is low. Thus, the total effect of pollution on probability of illness is $(24/100 - 0/100) - (0/100 - 0/100) = 0.24$. To compute the causal effect, hold behavior constant. Among the 20 children who face increased pollution but continue to play outdoors, the increased probability of illness is $(8/20 - 0/20) = 0.4$, whereas the change in probability of illness among the children who do not face increased pollution but continue to play outdoors is $(0/100 - 0/100) = 0$. Holding behavior constant, the causal effect of pollution is $0.4 - 0 = 0.4$. The effect of pollution on the probability of averting behavior may be computed by comparing the increased probability of staying indoors between those facing increased pollution $(80/100 - 0/100)$ and those facing no increased in pollution $(0/100 - 0/100)$ to obtain $0.8 - 0 = 0.8$. Finally, among those who face increased pollution, the probability of illness increases by $(16/80 - 0/80) = 0.2$ for those who engage in averting behavior by staying indoors and by $(8/20 - 0/20) = 0.4$ among those who do not engage in averting behavior. Averting behavior changes the probability of illness by $0.2 - 0.4 = -0.2$.

Table 8.2 Effects measured in the hypothetical natural experiment

Effect	Group A (more resistant to illness)	Group B (less resistant to illness)	Average effect in total population	Measured in total population if resistance to illness not controlled
Partial effect of pollution on probability of illness	0.20	0.40	0.30	0.24
Effect of averting on probability of illness	−0.20	−0.20	−0.20	−0.08
Effect of pollution on probability of averting	0.20	0.80	0.50	0.50
Total effect of pollution on probability of illness	0.16	0.24	0.20	0.20
Willingness to pay to avoid increased pollution	$20.00	$40.00	$30.00	$60.00

resistance to illness) is related both to health and to behavior, the partial effect of pollution on health is badly measured even if averting behavior is controlled.

The effect of averting behavior on the probability of illness is also badly measured if differential resistance to illness is unobserved. In Panel C of Table 8.1, among children who face high pollution on Day 2 but stay indoors, 16 of 100 become sick compared to none on Day 1. In contrast, 24 of the 100 children who face high ozone on Day 2 and play outdoors become sick, compared to none who were sick when facing low ozone levels on Day 1. Being unable to observe that the children are unequally prone to become ill, the effect of averting on the probability of illness appears to be $0.16 - 0.24 = -0.08$. This measure does not correspond to the true effect of averting behavior (-0.2) on average in the population or in either group. Thus, if an unobserved factor (such as resistance to illness) is related both to health and to behavior, the causal effect of averting behavior on health is badly measured.

Unless the errors in measuring the effects of pollution and behavior are coincidentally offsetting, willingness to pay will be badly measured if the differential resistance to illness is unobserved. Applying Eq. (8.3) to data in Panel C of Table 8.1, willingness to pay to avoid the increase in ambient pollution appears to be $\$20(0.24/0.08) = \$60$. This measure does not equal average willingness to pay in the total population ($30) or in either group ($20 in Group A or $40 in Group B). Thus, if an unobserved factor (such as resistance to illness) is related both to health and to behavior, the willingness to pay to reduce pollution is badly measured.

The problem of invalid measurement of the partial effects of pollution and of averting behavior on health arises because there is an uncontrolled factor that is related both to the health outcomes experienced and to the decision to avert. The partial effect of pollution is understated because the children who remain outdoors despite the increase in pollution are disproportionately drawn from the group of children least likely to become ill when pollution rises: 80% of the children who remain outdoors in high ozone conditions are members of Group A (those more resistant to illness). The effect of averting behavior on health is understated because those who avert are disproportionately drawn from those who are most likely to become ill when pollution rises: 80% of the children who avert are members of Group B (those less resistant to illness).

The presence of a "third factor" related to both health and behavior would be expected in almost any context in which averting behavior methods might be applied. Dealing with the resulting problem of identifying effects of the environment or of behavior on outcomes is central to valid application of averting behavior methods.

Further analysis of this example provides insight into one way of working around the problem of invalid measurement of the effects of pollution and of behavior arising from an unobserved third factor. Note that even if differential resistance to illness is unobserved, the effect of pollution on averting behavior $[(100/200 - 0/200) - (0/200 - 0/200) = 0.5]$ and the total effect of pollution on health $[(40/200 - 0/200) - (0/200 - 0/200) = 0.2]$ are measured validly using data in Panel C of Table 8.1 (assuming random assignment of pollution). The fact that the total effect of pollution and the effect of pollution on behavior are measured validly suggests that finding a way to rely on these measures, rather than on Eq. (8.3), to estimate willingness to pay might be a solution to the problem of an unobserved third factor for welfare measurement.

Proceeding along the lines suggested by Harrington and Portney (1987), multiply Eq. (8.1) by the negative of the unit cost of averting action, divide by the effect of averting behavior on health, and rearrange to obtain Eq. (8.4):

$$
\begin{aligned}
&-(\text{Unit cost of averting}) \times \left(\frac{\text{Partial effect of pollution on illness}}{\text{Effect of averting on illness}}\right) \\
&= -\left(\frac{\text{Unit cost of averting}}{\text{Effect of averting on illness}}\right) \times \begin{pmatrix}\text{Total effect}\\ \text{of pollution}\\ \text{on illness}\end{pmatrix} + (\text{Unit cost of averting}) \times \begin{pmatrix}\text{Effect of}\\ \text{pollution on}\\ \text{averting}\end{pmatrix}.
\end{aligned} \tag{8.4}
$$

Now substitute from Eqs. (8.2) and (8.3) to express the value of reduced pollution as

$$(\text{WTP for reduced pollution}) = \begin{pmatrix}\text{WTP for reduced}\\ \text{illness}\end{pmatrix} \times \begin{pmatrix}\text{Total effect of}\\ \text{pollution on illness}\end{pmatrix} + (\text{Unit cost of averting}) \times \begin{pmatrix}\text{Effect of pollution}\\ \text{on averting}\end{pmatrix}. \quad (8.5)$$

Based on data from Panel C of Table 8.1 and using the value of \$100 per day of avoided illness, Eq. (8.5) implies that the willingness to pay to avoid increased ozone concentrations is $\$100 \times 0.2 + \$20 \times 0.5 = \$30$. This value matches the average willingness to pay in the total population of children.

Notice that the first product on the right-hand side of Eq. (8.5) equals the valuation measure of the damage function approach, and the second product measures the effect of pollution on the costs of averting action. Equation (8.5) shows how the damage function approach misstates willingness to pay by ignoring the costs of averting behavior. If averting behavior increases with increased pollution, the damage function method understates willingness to pay.

The advantage of measuring willingness to pay for reduced pollution using Eq. (8.5) instead of Eq. (8.3) is that Eq. (8.5) avoids the need to measure the partial effect of pollution on health and the effect of averting behavior on health. These effects are difficult to measure accurately when there is an unobserved third factor related to health and to averting behavior. Equation (8.5) instead relies on the total effect of pollution on health and the effect of pollution on averting behavior. These effects are measured validly despite the presence of a third factor (assuming random assignment of pollution). A disadvantage of using Eq. (8.5) is that a measure of the willingness to pay for a change in health must be obtained from some source before the willingness to pay for reduced pollution can be measured. Thus, benefit estimation using Eq. (8.5) combines averting behavior and benefit transfer methods (see Chap. 11).

8.3 Theoretical Foundation of Averting Behavior Methods

This section describes the theoretical foundation of averting behavior methods, beginning with a presentation of a simple averting behavior model in Sect. 8.3.1. Section 8.3.2 derives measures of marginal willingness to pay for improvements in environmental quality and in outcomes of environmental exposure, and Sect. 8.3.3 describes the two methods used to estimate the value of nonmarginal changes in environmental quality based on averting behavior. Following this Sect. 8.4, presents a model of averting behavior under risk.

8.3.1 Theoretical Foundation

Averting behavior methods can be illustrated using a simple household production model (Courant and Porter 1981). Variations of this basic model have been used for many years in health economics and environmental economics.

An individual's preferences are represented by the quasi-concave utility function

$$U = U(x, h), \tag{8.6}$$

where x represents consumption of a private market good, h denotes consumption of a home-produced good, and $\partial U/\partial x > 0, \partial U/\partial h > 0$. The good h is the output of the household production function

$$h = f(I, q), \tag{8.7}$$

where I represents usage of a private good input, q denotes environmental quality, and $\partial f/\partial I > 0, \partial f/\partial q > 0$. The input I represents averting behavior because the individual can offset the effect of reduced environmental quality on the home-produced output using more I.[8]

Two variations on the specification in Eqs. (8.6) and (8.7) commonly arise in applied work. One variation concerns whether the environmental input is a "good" (e.g., environmental quality) or a "bad" (e.g., ambient pollution). When q represents a bad and the output h represents a good, $\partial f/\partial q < 0$.[9] The second variation occurs when the home-produced output h also is a bad (i.e., $\partial U/\partial h < 0$) rather than a good. When both h and q are "bads," assume $\partial U/\partial h < 0$, $\partial f/\partial q > 0, \partial f/\partial I < 0$.[10] In any variation of the model, however, the individual takes the amount of the environmental input q as given but chooses how much of input I to use. Thus, the outcome experienced (h) is endogenous because it is influenced by the individual's choices.

The budget constraint is

$$y = x + pI, \tag{8.8}$$

where the price of the market consumption good is normalized to unity, p denotes the relative price of the averting input, and y represents income. Two variations on

---

[8] In the model by Gerking and Stanley (1986), for example, h = "health capital," I = use of medical care, and q = ambient air quality.

[9] In Jakus (1994), for example, the home-produced output (h) is the quality of trees on an individual's property, the environmental input (q) is pest infestation, and the defensive behavior (I) is effort to control pests, such as spraying pesticide.

[10] For example, Richardson et al. (2012) estimated models in which h = symptoms of illness. Neidell (2004, 2009) and Moretti and Neidell (2011) considered situations where h represents emergency rooms visits or hospital admissions. Joyce et al. (1989) estimated health production functions where the outputs include h = probability of infant mortality. In all of these papers, the environmental inputs are measures of air pollution.

the budget constraint in Eq. (8.8) are common in applied work. One variation arises when it is important to account for the financial consequences of changes in environmental quality or home-produced output. A damage cost function $C(h)$ then is included in the expenditures on the right-hand side of Eq. (8.8), with $dC/dh < 0, d^2C/dh^2 \geq 0$. In health applications, for example, damage costs may be specified as the individual's costs of illness (e.g., $C(h) = M(h) + wG(h)$; Harrington and Portney 1987). The first component of the costs of illness is remedial medical expenses $M(h)$. The second component is time lost from market and nonmarket activities $G(h)$, valued at the wage rate w. Remedial medical expenses and time lost decline as health improves: $dM/dh < 0, dG/dh < 0$.

Another variation on the budget constraint arises when the allocation of time is a central feature of an averting behavior model. Accounting for the scarcity of time could be important because the output of the home production function is measured in units of time (Dickie 2005), use of the averting input requires time (Dickie and Gerking 1991b), or the averting behavior involves adjustments in time allocation such as spending less time outdoors (Neidell 2009). A time constraint is included in the model if the allocation of time is considered. See Harrington and Portney (1987) for details.

The averting behavior method is based on a model like the one summarized by Eqs. (8.6) through (8.8) in which an individual maximizes utility subject to a household production function and a budget constraint (and a time constraint when appropriate). Substituting the household production function into the utility function, the Lagrange function is

$$L = U[x, f(I, q)] + \lambda[y - x - pI], \tag{8.9}$$

where λ denotes the Lagrange multiplier. First-order necessary conditions are

$$\begin{aligned} &\partial U/\partial x - \lambda = 0, \\ &(\partial U/\partial h)\partial f/\partial I - \lambda p = 0, \\ &y - x - pI = 0. \end{aligned} \tag{8.10}$$

To interpret the equilibrium of the model, note that the first-order conditions imply that

$$\frac{\partial U/\partial h}{\partial U/\partial x} = \frac{\partial U/\partial h}{\lambda} = \frac{p}{\partial f/\partial I}. \tag{8.11}$$

According to Eq. (8.11), the individual allocates resources so that the marginal benefit of h equals the marginal cost of producing it. The marginal cost of h on the right-hand side of Eq. (8.11) equals the price of the averting input relative to its marginal product. The marginal benefit of h on the left-hand side of Eq. (8.11) is reflected in the marginal rate of substitution of h for x, measuring the amount of private consumption x the individual is willing to give up for another unit of

h. Because the price of x is normalized to unity, this marginal rate of substitution also measures the amount of income the individual is willing to pay for an additional unit of h. Because the Lagrange multiplier λ equals the marginal utility of income at the maximum, this marginal willingness to pay for h equals the monetized marginal utility of h, $(\partial U/\partial h)/\lambda$. Eq. (8.11) represents the analytical derivation of Eq. (8.2).

The second-order condition sufficient for a constrained maximum is

$$\Delta = -p^2\left(\frac{\partial^2 U}{\partial x^2}\right) + 2p\left(\frac{\partial f}{\partial I}\right)\left(\frac{\partial^2 U}{\partial h \partial x}\right) - \left(\frac{\partial f}{\partial I}\right)^2\left(\frac{\partial^2 U}{\partial h^2}\right) - \left(\frac{\partial U}{\partial h}\right)\left(\frac{\partial^2 f}{\partial I^2}\right) > 0. \tag{8.12}$$

Assuming the second-order condition is satisfied, the first-order conditions may be solved for the optimal values of consumption x, averting input I and λ as functions of the relative price of defensive behavior p, environmental quality q and income y. The solution yields the Marshallian demands shown in Eq. (8.13).

$$\begin{aligned} x^* &= x(p,q,y), \\ I^* &= I(p,q,y). \end{aligned} \tag{8.13}$$

The reduced-form demand function for I in Eq. (8.13) shows how the utility-maximizing individual chooses averting action.

Substitute the demand for averting input into the production function to obtain the utility-maximizing amount of home-produced output, or the "demand" for h:

$$h^* = f[I^*(p,q,y),q] = h(p,q,y). \tag{8.14}$$

Equation (8.14) is a reduced-form equation for the home-produced output and is an example of a damage function. Finally, substitute the optimal values of x and h into the utility function to obtain the indirect utility function:

$$v(p,q,y) = U[x(p,q,y),h(p,q,y)] = U[x(p,q,y),f(I(p,q,y),q)]. \tag{8.15}$$

As an example, suppose the household production function takes the Cobb–Douglas form $h = GI^{\gamma_I}q^{\gamma_q}$, and the utility function takes the quasi-linear form $U = x + uh^{\alpha}$, where $\gamma_I > 0$, $\gamma_q > 0$, $\alpha > 0$, $u > 0$. Let $\beta_I = \alpha\gamma_I$ and $\beta_q = \alpha\gamma_q$. The second-order sufficient condition is satisfied if $\beta_I < 1$. In this specification, the marginal cost in Eq. (8.11) equals $pI^*/\gamma_I h^*$. Also,

$$\begin{aligned} I^* &= I(p,q,y) = \left(G^{\alpha}\beta_I p^{-1} q^{\beta_q} u\right)^{1/(1-\beta_I)}, \\ h^* &= h(p,q,y) = \left(G\beta_I^{\gamma_I} p^{-\gamma_I} q^{\gamma_q} u^{\gamma_I}\right)^{1/(1-\beta_I)}, \\ v(p,q,y) &= y + ((1-\beta_I)/\beta_I)\left(G^{\alpha}\beta_I p^{-\beta_I} q^{\beta_q} u\right)^{1/(1-\beta_I)}. \end{aligned} \tag{8.16}$$

8.3.2 Marginal Willingness to Pay for Improved Environmental Quality

The marginal willingness to pay for improved environmental quality is obtained from the indirect utility function as $(\partial v/\partial q)/(\partial v/\partial y)$ (see Chap. 2). Applying the envelope theorem and using the Lagrange function, $\partial v/\partial q = (\partial U/\partial h)(\partial f/\partial q)$ and $\partial v/\partial y = \lambda$. Using Eq. (8.11), marginal willingness to pay for improved environmental quality is given by

$$\frac{\partial v/\partial q}{\partial v/\partial y} = \left(\frac{\partial U/\partial h}{\lambda}\right)(\partial f/\partial q) = \left(\frac{\partial U/\partial h}{\partial U/\partial x}\right)(\partial f/\partial q) = p\frac{\partial f/\partial q}{\partial f/\partial I}. \tag{8.17}$$

According to Eq. (8.17), marginal willingness to pay for improved environmental quality equals the marginal value of a unit of the home-produced output h, $((\partial U/\partial h)/\lambda)$, weighted by the marginal effect of environmental quality on h, $(\partial f/\partial q)$. Based on maximizing behavior, this value in turn equates to the marginal cost of h, $p/(\partial f/\partial I)$, weighted by the marginal effect of environmental quality on h. Equation (8.17) presents the analytical derivation of Eq. (8.3) and is the same result as Eq. (2.23) in Chap. 2. In the example with a Cobb–Douglas production function, marginal willingness to pay equals $(\gamma_q/\gamma_I)(I^*/q)$.

Ten features of the marginal willingness-to-pay expression in Eq. (8.17) warrant further discussion.

(1) Eq. (8.17) indicates that the individual is willing to pay more to improve environmental quality when defensive behavior is expensive (large p) or ineffective (small $\partial f/\partial I$), or when environmental quality has a substantial impact on the home-produced output (large $\partial f/\partial q$). Put differently, marginal willingness to pay to improve environmental quality depends on the individual's opportunities to substitute averting behavior for environmental quality as measured by the marginal rate of technical substitution between the defensive input and environmental quality: $((\partial f/\partial q)/(\partial f/\partial I))$ (Courant and Porter 1981).

(2) No utility terms appear on the right-hand side of Eq. (8.17). As pointed out by Courant and Porter (1981), marginal willingness to pay may be inferred from the technology (the household production function, which in principle is an objective relationship) without information about tastes (the subjective utility function). Many researchers, beginning with Gerking and Stanley (1986), have used this insight to infer willingness to pay from estimated parameters of the household production technology.

(3) As implied by Eq. (2.23) in Chap. 2, the marginal willingness to pay expression in Eq. (8.17) can be interpreted as a virtual price of environmental quality.

(4) The willingness-to-pay expression in Eq. (8.17) does not fit the damage function approach typically applied to conduct benefit-cost analyses for environmental regulations. In the health context, for example, economists usually

are not asked to value a change in environmental quality directly as in Eq. (8.17), but instead to value a given change in health status as in Eq. (8.11). The health change to be valued and its link to environmental quality are determined by damage function estimates from epidemiology or toxicology. For example, health scientists estimate how environmental quality affects days of symptoms or risk of death. Economists estimate a value for reducing days of symptoms or risk of death. Policy analysts put the two estimates together to estimate the value of improved environmental quality. The demands of environmental policy analysis often call for measuring the value of the change in health in Eq. (8.11) rather than the value of improved environmental quality in Eq. (8.17).

(5) Marginal willingness to pay for improved environmental quality equals the savings in defensive expenditures (expenditures on averting behavior) that would hold the home-produced output constant despite improved environmental quality. In the case of one averting input, determine the amount of averting input required to produce a given level of home-produced output (h^0) at a specified level of environmental quality (q^0). Set up the identity $h^0 - f(I, q^0) \equiv 0$ and use the implicit function theorem to define the required amount of defensive input $I^D = I^D(q^0, h^0)$. According to the implicit function rule,

$$\partial I^D/\partial q = -(\partial f/\partial q/\partial f/\partial I) < 0. \tag{8.18}$$

Equation (8.18) indicates that the marginal rate of technical substitution $(\partial f/\partial q/\partial f/\partial I)$ gives the decrease in I that holds h constant when q increases. Now define (Bartik 1988) $D(p, q^0, h^0)$ as the function giving the required defensive expenditure to produce output level h^0 when environmental quality is q^0 and the price of the defensive input is p The defensive expenditure function is

$$D(p, q^0, h^0) = pI^D(q^0, h^0). \tag{8.19}$$

The change in defensive expenditure that holds output constant when environmental quality changes is

$$\partial D/\partial q = -p(\partial f/\partial q/\partial f/\partial I) < 0. \tag{8.20}$$

Thus $p((\partial f/\partial q)/(\partial f/\partial I))$ measures the savings in expenditures on the averting input that would hold constant the home-produced output despite an improvement in environmental quality. As shown in Eq. (8.17), this equates to marginal willingness to pay.[11]

---

[11]If there is more than one averting input (e.g., I_1 and I_2 with relative prices p_1 and p_2), one finds the quantities of inputs that minimize the cost of producing a given output level at a given level of environmental quality. The cost-minimizing demands are $I_j(p_1, p_2, q, h)$, $j = 1, 2$. The

(6) The actual or observed change in defensive expenditures does not in general equal willingness to pay. The change in averting input chosen by a utility-maximizing individual is obtained from the Marshallian demand function in Eq. (8.13) as

$$\frac{\partial I^*}{\partial q} = \frac{-p\left(\frac{\partial f}{\partial q}\right)\left(\frac{\partial^2 U}{\partial x \partial h}\right) + \left(\frac{\partial f}{\partial q}\right)\left(\frac{\partial f}{\partial I}\right)\left(\frac{\partial^2 U}{\partial h^2}\right) + \left(\frac{\partial^2 f}{\partial I \partial q}\right)}{\Delta}. \tag{8.21}$$

The sign of $\partial I^*/\partial q$ is indeterminate, as is the sign of the corresponding change in defensive expenditure, $p(\partial I^*/\partial q)$.[12] In contrast, the change in defensive expenditures required to hold output constant when environmental quality improves is negative. The change in defensive expenditure observed among utility-maximizing individuals may not even have the same sign as the change that holds output constant. Thus the observed savings in defensive expenditures is a poor approximation of willingness to pay (Courant and Porter 1981).[13]

In general, the relationship between the observed change in defensive expenditures and the change that holds output constant is governed by

$$\partial I^*/\partial q = \partial I^D/\partial q + \left(\partial I^D/\partial h\right)(\partial h^*/\partial q). \tag{8.22}$$

This equation indicates that when environmental quality improves, there is an effect on the demand for defensive input arising from the marginal rate of technical substitution, because less I is needed to produce a given level of h when q is larger $(\partial I^D/\partial q < 0)$. But the environmental change may affect the utility-maximizing quantity of output, as indicated by $\partial h^*/\partial q$. A change in the desired level of h requires the individual to adjust the averting input by $\partial I^D/\partial h > 0$.

(Footnote 11 continued)

defensive expenditure function is the minimum cost function $D(p_1, p_2, q, h) = p_1 I_2(p_1, p_2, q, h) + p_2 I_2(p_1, p_2, q, h)$..

[12]Even the compensated, or utility constant, effect of environmental quality on the demand for a defensive input is indeterminate in sign. The Marshallian effect will be negative if the defensive input is more productive when environmental quality is low, the marginal utility of market consumption is greater when home-produced output is higher, and there is diminishing marginal utility in home-produced output.

[13]With a Cobb–Douglas production function, $I^D(h, q) = \left(G^{-1} h q^{-\gamma_q}\right)^{1/\gamma_I}$, and the change in defensive expenditure that holds health constant when environmental quality improves is $\partial D/\partial q = -\left(D\gamma_q/q\gamma_I\right) = -p\left(\gamma_q/\gamma_I\right)(I/q) < 0$. With a quasi-linear utility function, the utility-maximizing change in defensive expenditure is $p(\partial I^*/\partial q) = \left(\beta_q/(1-\beta_I)\right)(D/q) > 0$. Thus, the chosen change in defensive expenditure does not even take the same sign as the change that holds output constant. As another example, if both the production function and the utility function take the Cobb–Douglas form, the utility-maximizing change in defensive expenditures is zero, whereas the change that holds output constant remains negative and takes the same form as in the example considered in the text.

(7) A seventh important feature of the willingness-to-pay expression in Eq. (8.17) is that it depends on the partial effect of environmental quality on the home-produced output with defensive behavior held constant, $(\partial f/\partial q)$, rather than the total effect (Gerking and Stanley 1986). The divergence between partial and total effects is fundamental because the distinguishing feature of an averting behavior model is optimization subject to a household production function that summarizes opportunities to defend against environmental hazards.

Consider the effect of a change in water quality on health, h. Holding I and everything else constant, the partial effect of water contamination on health is $(\partial f/\partial q)$. If water quality is randomly assigned, this partial effect represents the causal effect of water quality on health. Although a measure of this effect is needed to estimate marginal willingness to pay in Eq. (8.17), the partial effect is not the effect observed when people take averting action.

Suppose that people react to changes in water quality by adjusting their usage of bottled water I. Accounting for the change in behavior, the total effect of water quality on health is

$$\mathrm{d}h/\mathrm{d}q = [\partial f/\partial q + (\partial f/\partial I)(\partial I^*/\partial q)] = \partial h^*/\partial q. \tag{8.23}$$

Equation (8.23) presents the analytical derivation of Eq. (8.1) and represents the effect of environmental quality on health observed in a population of people who adjust behavior in response to environmental conditions. Equation (8.23) shows how the total effect of q on h consists of the partial effect, $(\partial f/\partial q)$, plus an indirect or behavioral effect. The indirect effect traces the impact of a change in environmental quality from the adjustment in defensive behavior $(\partial I^*/\partial q)$ through the effect of behavior on health $(\partial f/\partial I)$. The sign of $\partial I^*/\partial q$ is indeterminate (Eq. (8.21)). Thus, the total effect of a change in environmental quality may be greater than, equal to, or less than the partial effect, and the two effects need not even have the same sign.[14]

Ignoring averting behavior amounts to using the total effect of environmental quality instead of the partial effect when estimating willingness to pay, thus leading to a misstatement, and likely understatement, of the benefits of reducing pollution. Avoiding this problem requires disentangling the partial effect of environmental quality from the behavioral response. Furthermore, Eq. (8.17) also requires identifying the effect of averting behavior on health, $\partial f/\partial I$. Identifying effects of

[14]In the example with Cobb–Douglas production and quasi-linear utility, $\partial I^*/\partial q > 0$, as discussed above in connection with the change in defensive expenditures. Thus the total effect exceeds the partial effect, $\partial h^*/\partial q = (\partial f/\partial q)/(1-\beta_I) > (\partial f/\partial q) > 0$. Alternatively, if both the production function and the utility function take the Cobb–Douglas form, $\partial I^*/\partial q = 0$ and the partial and total effects coincide. Available empirical evidence suggests, in contrast to these two examples, that in most cases considered, $\partial I^*/\partial q < 0$ and the total effect of changes in environmental quality will be less than the partial effect. In any case, the two effects are not identical, and it is the partial effect that is needed to estimate marginal willingness to pay in Eq. (8.17).

environmental quality and of averting behavior on health, however, is quite challenging, as discussed in Sect. 8.5.

(8) As demonstrated by Harrington and Portney (1987), the need to identify partial effects of environmental quality and defensive behavior can be avoided if one has prior knowledge of the marginal value of health $((\partial U/\partial h)/\lambda)$. Solve Eq. (8.23) for the marginal rate of technical substitution, multiply by the price of averting input, and substitute from Eqs. (8.11) and (8.17) to obtain

$$\frac{\partial v/\partial q}{\partial v/\partial y} = p\frac{\partial f/\partial q}{\partial f/\partial I} = \left(\frac{\partial U/\partial h}{\lambda}\right)\left(\frac{\mathrm{d}h}{\mathrm{d}q}\right) - p\left(\frac{\partial I^*}{\partial q}\right). \tag{8.24}$$

According to Eq. (8.24), which gives the analytical derivation of Eq. (8.5), marginal willingness to pay to improve environmental quality can be measured without requiring a measure of the partial effect of environmental quality, holding behavior constant. Instead, one estimates the total effect of environmental quality, $\mathrm{d}h/\mathrm{d}q$. Furthermore, a measure of the effect of averting behavior on health, $\partial f/\partial I$, is not required. A disadvantage of this approach, on the other hand, is that Eq. (8.24) includes utility terms that are difficult to quantify. Some type of benefit transfer is needed to apply a value for the marginal value of health in Eq. (8.24) (see Chap. 11).

(9) Eq. (8.24) indicates that the damage function approach to valuation—applying a unit value for health $((\partial U/\partial h)/\lambda)$ to the total effect of environmental quality $(\mathrm{d}h/\mathrm{d}q)$—misstates willingness to pay by ignoring the costs of averting behavior $(p(\partial I^*/\partial q))$. If averting behavior declines when environmental quality improves, $\partial I^*/\partial q < 0$, then the damage function method understates marginal willingness to pay for improved environmental quality.

(10) Finally, the welfare measures in Eqs. (8.11) and (8.17) may be compared to corresponding measures of damage costs (Berger et al. 1987; Harrington and Portney 1987). If $C(h)$ represents damage costs, then the damage cost measure of the value of a marginal change in home-produced output equals $-\mathrm{d}C/\mathrm{d}h$. As discussed in Sect. 8.3.1, damage costs change the budget constraint to $y = x + pI + C(h)$. In the presence of damage costs, marginal willingness to pay for an improvement in environmental quality is

$$\frac{\partial v/\partial q}{\partial v/\partial y} = \left[\frac{\partial U/\partial h}{\lambda} - \mathrm{d}C/\mathrm{d}h\right](\partial f/\partial q) = p\frac{\partial f/\partial q}{\partial f/\partial I}. \tag{8.25}$$

Thus, marginal willingness to pay exceeds the absolute value of the reduction in damage costs by the amount $(\partial U/\partial h)/\lambda$. In other words, damage costs fall short of willingness to pay because they focus on resource costs while ignoring the utility value of a change in home-produced output. In the health context damage costs are measured by the costs of illness. The marginal value of a given reduction in illness

$dh > 0$ exceeds the cost of illness, because the cost of illness does not account for the utility value of health or the "pain and suffering" of illness.[15]

8.3.3 Valuing Nonmarginal Changes in Pollution

Two approaches have been taken to use information on averting behavior to value nonmarginal changes in environmental quality. The Bockstael–McConnell method (Bockstael and McConnell 1983) estimates the exact compensating surplus and the Bartik (1988) approach estimates a lower bound on the compensating surplus.

The compensating surplus for a nonmarginal change in environmental quality may be derived from the expenditure function $e(p, q, U)$ that measures the minimum expenditure necessary to achieve the utility level U when the price of the defensive input is p and environmental quality is q (see Chap. 2). Suppose that $U^0 = v(p^0, q^0, y)$. As discussed in Chap. 2, the compensating surplus for a change in environmental quality from q^0 to q^1 equals

$$CS = e\left(p^0, q^0, U^0\right) - e\left(p^0, q^1, U^0\right). \tag{8.26}$$

In the example with Cobb–Douglas production and quasi-linear utility, the expenditure function is $e(p, q, U) = U - ((1 - \beta_I)/\beta_I)\left(\beta_I G^\alpha p^{-\beta_I} q^{\beta_q} u\right)^{1/(1-\beta_I)}$, and the compensating surplus is $CS = ((1 - \beta_I)/\beta_I)\left(\beta_I G^\alpha p^{-\beta_I} u\right)^{1/(1-\beta_I)}\left[\left(q^1\right)^{\beta_q/1-\beta_I} - \left(q^0\right)^{\beta_q/1-\beta_I}\right]$.

Bockstael and McConnell (1983) developed a way to measure the compensating surplus in Eq. (8.26) using the area behind the compensated (Hicksian) demand curve for an averting input. Their approach requires that the averting action, I, be a necessary or essential input in the production of the output, h, and that environmental quality, q, be weakly complementary to the output. The Bockstael–McConnell method generalizes the analysis of weak complementarity that was presented in Chap. 2 to incorporate household production.[16]

---

[15]The relationship between estimates of damage costs and willingness to pay, however, is less clear cut when considering a change in environmental quality rather than a change in health or other home-produced output. This occurs because damage cost estimates normally rely on the total effect of environmental quality dh/dq rather than the partial effect used to estimate marginal willingness to pay in Eq. (8.17). The relative magnitude of partial and total effects is theoretically ambiguous, but prior empirical evidence suggests that defensive behavior increases when environmental quality diminishes, causing the partial effect to exceed the total effect in absolute value. This outcome leads to the presumption that willingness to pay for improved environmental quality exceeds the savings in damage costs associated with improved environmental quality.

[16]The Bockstael–McConnell model is more general than the one considered here in that it allows q to enter the utility function directly (as opposed to being only an input into a household production function), and allows for multiple home-produced goods.

Define the compensated demand function for the averting input, I, as $I^h(p,q,U^0) = \partial e(p,q,U^0)/\partial p$. Denote the area behind the compensated demand curve (the area below the curve and above the prevailing price, $p = p^0$) for a given level of environmental quality, $q = q^t$, as

$$A^t = \int_{p^0}^{p^c} I^h(p,q^t,U^0)\mathrm{d}p = e(p^c,q^t,U^0) - e(p^0,q^t,U^0). \tag{8.27}$$

The change in area behind the compensated demand curve when q changes from q^0 to q^1 is

$$A^1 - A^0 = CS + \left[e\left(p^c,q^1,U^0\right) - e\left(p^c,q^0,U^0\right)\right]. \tag{8.28}$$

In Eqs. (8.27) and (8.28), p^c denotes the "choke price" that drives compensated demand to zero so that $I^h(p^c,q,U) = 0$. Because the compensated demand shifts with changes in environmental quality or utility, the choke price is in general a function of q and U: $p^c(q,U)$.

Equation (8.28) indicates that the change in area behind the compensated demand for an averting input equals compensating surplus if the term in brackets equals zero. The term in brackets measures the change in expenditures needed to attain utility U^0 when environmental quality changes from q^0 to q^1 and the price of the averting input equals the choke price. The change in expenditures equals[17]

$$\mathrm{d}e\left(p^c,q,U^0\right)/\mathrm{d}q = \mu[(\partial U/\partial h)(\partial f/\partial q) - (\partial U/\partial q)], \tag{8.29}$$

where μ denotes the Lagrange multiplier from the expenditure minimization problem.

The change in expenditures in Eq. (8.29) equals zero, and consequently, the change in area behind the Hicksian demand curve equals the compensating surplus, if two conditions hold: (1) If I is a necessary input in the production of h, then $f(0,q) \equiv 0$ for all values of q. This condition implies that $\partial f/\partial q \equiv 0$ when $I = 0$, at the choke price. (2) If environmental quality is weakly complementary to h (see

[17]The change in expenditures equals

$$\begin{aligned}\mathrm{d}e\left(p^c,q,U^0\right)/\mathrm{d}q =& \left(\partial e\left(p^c,q,U^0\right)/\partial p\right)(\partial p^c/\partial q) - \left(\partial e\left(p^c,q,U^0\right)/\partial q\right)\\ =& I^h\left(p^c,q,U^0\right)(\partial p^c/\partial q) - (\mu\partial U/\partial h)(\partial f/\partial q) - \mu(\partial U/\partial q)\\ =& 0(\partial p^c/\partial q) - (\mu\partial U/\partial h)(\partial f/\partial q) - \mu(\partial U/\partial q)\end{aligned}$$

.

Chap. 2), then $\partial U/\partial q \equiv 0$ when $h = 0$. Note that this condition holds automatically if q does not enter the utility function.[18]

An advantage of this approach is the ability to estimate welfare changes using the demand for only one input, as opposed to estimating the household production function. The simplification comes at the expense of assuming the input is necessary, an assumption that could be difficult to justify in many circumstances (Smith 1991). Dickie and Gerking (1991b) applied the Bockstael–McConnell method to estimate benefits of reduced ozone pollution, and Agee and Crocker (1996) applied it to estimate benefits of lowered lead burdens in children.

In the example model, I is a necessary input because $f(0, q) \equiv 0$, and h and q are weakly complementary because q does not enter the utility function directly. Thus, sufficient conditions for applying the Bockstael–McConnell approach are met. The compensated demand for the defensive input is $I^h(p, q, U) = \partial e(p, q, U))/\partial p = \left(\beta_I G^\alpha p^{-1} q^{\beta_q} u\right)^{1/1-\beta_I}$.[19] This demand function has no finite choke price because $I^h(p, q, U) > 0$ for all values of $p > 0$. Nonetheless, the Bockstael–McConnell method is applicable because the integral of the demand function, $A = \left(\beta_I G^\alpha q^{\beta_q} u\right)^{1/1-\beta_I} \int_{p^0}^{\infty} p^{-1/1-\beta_I} \mathrm{d}p$, converges for $0 < \beta_I < 1$.[20] The change in area when q changes from q^0 to q^1 is $A^1 - A^0 = ((1 - \beta_I)/\beta_I) \left(\beta_I G^\alpha p^{-\beta_I} u\right)^{1/(1-\beta_I)} \left[(q^1)^{\beta_q/1-\beta_I} - (q^0)^{\beta_q/1-\beta_I}\right] = CS$.

An alternative approach is to use the change in defensive expenditures as a bound on the compensating surplus. Recall that the willingness to pay for a marginal improvement in environmental quality equals the savings in defensive expenditures that holds home-produced output constant. Bartik (1988) extended this result to show that the willingness to pay for a nonmarginal improvement in environmental quality is bounded from below by the savings in defensive expenditures that maintains the initial level of the home-produced output.

Using the defensive expenditure function in Eq. (8.19), the savings in defensive expenditures that maintains household output level h^0 when environmental quality improves from q^0 to q^1 is

$$D(p, q^0, h^0) - D(p, q^1, h^0) = p\left[I^D\left(p, q^0, h^0\right) - I^D\left(p, q^1, h^0\right)\right]. \qquad (8.30)$$

Bartik (1988) showed that the measure in Eq. (8.30) is a lower bound on the compensating surplus.

---

[18]Bockstael and McConnell (1983) explained how these two conditions generalize to the case in which there are multiple home-produced outputs and discussed approximating the compensating surplus using the ordinary (Marshallian) demand curve for a defensive input.

[19]The compensated and ordinary demand functions are identical because of the quasi-linear form of the utility function.

[20]Recall that $\beta_I = \alpha\gamma_I > 0$ because h is a good $(\alpha > 0)$ and q is a good $(\gamma_I > 0)$. Also, $\beta_I < 1$ to meet second-order sufficient conditions.

The change in defensive expenditure in Eq. (8.30) is computed holding the home-produced output constant. Bartik's (1988) bounding result does not necessarily apply to the observed change in defensive expenditures, because the individual presumably would adjust h in response to the change in q (see Eq. (8.22)). But if the individual chooses to increase h as q increases, the observed change in defensive expenditures is also a lower bound on compensating surplus, albeit a less-accurate bound than the change in expenditures holding output constant. As a practical matter, observed defensive expenditures are much easier to calculate because they do not require estimation (or prior knowledge) of the defensive expenditure function.[21]

Murdoch and Thayer (1990) applied Bartik's (1988) analysis to estimate the change in expenditures on sunscreen lotion that would hold the incidence of nonmelanoma skin cancer constant despite a change in ultraviolet radiation exposure due to depletion of stratospheric ozone. Most applications of Bartik's analysis, however, use observed defensive expenditures and involve avoidance of contaminated water supplies by some combination of purchasing bottled water, boiling or treating water at home, hauling water from an alternate source, or installing water filtration systems (for example, see, Harrington et al., 1989).

8.4 Averting Behavior Under Risk

Many applications of averting behavior methods involve situations in which the individual does not know with certainty the outcomes that may be caused by environmental conditions or by averting action. To introduce risk, assume the individual must make decisions before knowing which of two possible "states of the world" will occur. The states of the world describe alternative outcomes that might be experienced. For example, in a model used to value reduced risk of illness, the individual experiences good health in one state of the world and ill health in the other state.

A simple one-period model focused on mortality risk illustrates how the averting behavior method is applied to risk. In the model, the individual survives in one state of the world, but dies in the other. This model supports estimation of the willingness to pay to reduce mortality risk and the value of a statistical life. Utility is $U_a(x)$ if the individual is alive (a), and $U_d(x)$ if he dies (d), where x represents resources available for consumption.[22] The utility gain from survival is $U_a(x) - U_d(x) > 0$. The marginal utility of consumption is $U'_a(x) > 0$ if the individual

---

[21]In the example model, $D(p, q^0, h^0) - D(p, q^1, h^0) = \left[\beta_I G^\alpha p^{-\beta_I} (q^0)^{\gamma_q/\gamma_I} u\right]^{1/1-\beta_I} \left[(q^0)^{-\gamma_q/\gamma_I} - (q^1)^{-\gamma_q/\gamma_I}\right]$. The observed change in defensive expenditures is $pI(p, q^0, y) - pI(p, q^1, y) = \left[\beta_I G^\alpha p^{-\beta_I} u\right]^{1/1-\beta_I} \left[(q^0)^{\beta_q/1-\beta_I} - (q^1)^{\beta_q/1-\beta_I}\right]$.

[22]Utility in death is often considered "bequest utility." The key assumption is that $U_a(x) > U_d(x)$.

survives and $U'_d(x) \geq 0$ if the individual dies. The probability of survival is denoted s, and expected utility is given by $EU = sU_a(x) + (1-s)U_d(x)$. The expected marginal utility of consumption, denoted $E[MUC]$ below, is $E[MUC] = sU'_a(x) + (1-s)U'_d(x) > 0$.

Now consider the amount of money the individual would be willing to pay to increase the probability of survival by a small amount in advance of knowing which state of the world will occur. First, consider an initial survival probability s^0 and the resulting expected utility $EU^0 = s^0 U_a(x) + (1-s^0)U_d(x)$. The willingness to pay W to improve the chance of survival to s^1 is defined implicitly by the identity

$$s^1 U_a(x-W) + \left(1-s^1\right)U_d(x-W) - EU^0 \equiv 0. \tag{8.31}$$

Provided that identity (8.31) implicitly defines W as a function of s, given values of x and EU^0, obtain the marginal value of reduced mortality risk using the implicit function rule as[23]

$$-\frac{\mathrm{d}W}{\mathrm{d}s} = \frac{U_a(x) - U_d(x)}{E[MUC]}. \tag{8.32}$$

The marginal value in Eq. (8.32) gives the amount of consumption the individual is willing to sacrifice to increase survival probability by a small amount. According to Eq. (8.32), the individual will give up more consumption in exchange for increased probability of survival if (1) the utility gain from survival is larger or (2) the expected marginal utility of consumption is smaller.[24]

Now consider how information on averting behavior might be used to measure the value of reduced mortality risk (Blomquist 1979). Suppose that the probability of survival depends on ambient air quality and defensive behavior, $s = s(I, q)$, where $\partial s/\partial I > 0, \partial s/\partial q > 0$. The budget constraint is $y = x + pI$. Use the budget constraint to substitute for x in the state-dependent utility functions and choose I to maximize $EU = s(I,q)U_a(y-pI) + [1 - s(I,q)]U_d(y-pI)$.

The first-order necessary condition for maximization of expected utility,[25] $(\partial s/\partial I)[U_a(x) - U_d(x)] - p\left[s(I,q)U'_a(x) + (1 - s(I,q))U'_d(x)\right] = 0$, implies that

[23]To see this, differentiate the identity (8.37) with respect to W and S to obtain $(U_a - U_d)\mathrm{d}s - \left(sU'_a + (1-s)U'_d\right)\mathrm{d}W$. Set the differential equal to zero to hold expected utility constant and solve for $\mathrm{d}W/\mathrm{d}s$.

[24]The value of statistical life is determined from Eq. (8.32) by applying the change in survival probability to a population sufficiently large to be expected to save one life and computing the amount the population would be willing to pay. For example, suppose the small change in survival probability evaluated in Eq. (8.31) is $\mathrm{d}s = 10^{-5}$. If a population of 10,000 persons each experienced this change in survival probability, one would expect the number of deaths in the population to decline by one. Thus, multiplying the value of reduced risk in Eq. (8.32) by 10,000 would yield the value of statistical life.

[25]The sufficient condition $-(\partial^2 s/\partial I^2)(U_a - U_d) + 2p(\partial s/\partial I)\left(U'_a - U'_d\right) + p^2 (\mathrm{d}E[MUC]/\mathrm{d}x) < 0$ holds if there is (1) diminishing marginal productivity of defensive action,

$$\frac{U_a(x) - U_d(x)}{E[MUC]} = \frac{p}{(\partial s/\partial I)}. \tag{8.33}$$

According to Eq. (8.33), the individual chooses averting action so that the marginal benefit of a reduction in probability of death equals the marginal cost of reducing the probability. Equation (8.33) is the version of Eqs. (8.2) and (8.11) that applies in the two-state model of risk. As in the model under certainty, willingness to pay under risk may be determined from the price and marginal product of an averting action. Furthermore, the marginal value of a small improvement in environmental quality is

$$\frac{U_a(x) - U_d(x)}{E[MUC]}(\partial s/\partial q) = p\frac{(\partial s/\partial q)}{(\partial s/\partial I)}. \tag{8.34}$$

Equation (8.34) is the version of Eqs. (8.3) and (8.17) arising in the two-state model with risk.[26]

8.5 Obstacles to Valid Welfare Measurement with Averting Behavior Methods

At least four key obstacles impede valid welfare measurement using averting behavior methods: (1) joint production, (2) badly measured prices of averting actions, (3) difficulty of identifying the partial effects of averting behavior and of environmental conditions, and (4) difficulty of identifying the total effect of environmental conditions. These challenges are discussed in turn in Sects. 8.5.1 through 8.5.4.

8.5.1 *Joint Production*

Joint production occurs when the effect of an averting behavior on individual welfare does not operate through a single outcome such as health or probability of survival, but instead affects more than one home-produced output or enters the

(Footnote 25 continued)

(2) a higher marginal utility of consumption when alive than when dead, and (3) financial risk loving in neither state of the world $\left(U_a''(x), U_d''(x) \leq 0\right)$.

[26]Blomquist (2004) reviewed pertinent conceptual and empirical issues involved in applying averting behavior models to estimate the value of reduced risk of death and provides citations to the literature. Results for models more complex than the two-state model were considered by Shogren and Crocker (1991), Quiggin (1992), and Bresnahan and Dickie (1995).

utility function directly. Joint production is pervasive in the household production setting (Pollak and Wachter 1975). For example, changing the source of drinking water may affect the taste, odor, appearance, convenience, or other attributes of water, as well as reducing exposure to contaminants. Staying indoors to avoid air pollution may reduce the incidence of respiratory illness while also affecting the enjoyment of leisure time. Purchases of safety goods, such as smoke detectors and bicycle helmets, jointly reduce at least two risks, risk of injury as well as risk of death.

Suppose that the averting input, I, enters the utility function directly so that $U = U(x, h, I)$. In this case, marginal willingness to pay to improve environmental quality is not given by Eq. (8.17) but by

$$\frac{\partial v/\partial q}{\partial v/\partial y} = \frac{\partial U/\partial h}{\lambda}(\partial f/\partial q) = \left(p - \frac{\partial U/\partial I}{\lambda}\right)\left(\frac{\partial f/\partial q}{\partial f/\partial I}\right). \tag{8.35}$$

Thus, marginal willingness to pay cannot be determined using information on the technology (the marginal rate of technical substitution between environmental quality and the defensive input) alone. Information on tastes (the marginal utility of the defensive input and of income) also enters the welfare expression. If Eq. (8.17) is used to measure marginal willingness to pay, the true value given by Eq. (8.35) is overstated if the averting input provides joint benefits $(\partial U/\partial I > 0)$. Intuitively, part of the expenditure on the input I is attributable to the joint benefit. The monetized value of the joint benefit $((\partial U/\partial I)/\lambda)$ must be deducted from the marginal expenditure p to isolate the part of the expenditure on I that is attributable to mitigation. Conversely, Eq. (8.17) understates willingness to pay if the averting input produces joint costs $(\partial U/\partial I < 0)$, because the market price p understates the full cost of the input by ignoring the monetized value of the disutility.

As a practical matter, the "solution" to joint production usually involves asserting an intuitive argument for the sign of $\partial U/\partial I$ to establish whether the welfare measure in Eq. (8.17) is an upper or lower bound on the true welfare measure in Eq. (8.35).[27]

The situation is somewhat similar if environmental quality indivisibly enters multiple household production functions or enters the utility function directly. If q enters the utility function, marginal willingness to pay for improved environmental quality equals

$$\frac{\partial v/\partial q}{\partial v/\partial y} = \frac{\partial U/\partial h}{\lambda}(\partial f/\partial q) + \frac{\partial U/\partial q}{\lambda} = p\left(\frac{\partial f/\partial q}{\partial f/\partial I}\right) + \frac{\partial U/\partial q}{\lambda}. \tag{8.36}$$

[27]Alternative approaches include the following: Blomquist (1979) applied a clever empirical strategy to attack joint production; Hori (1975) derived theoretical conditions, implemented empirically by Dickie and Gerking (1991a), that allow the problem to be overcome; Dardis (1980) simulated willingness to pay under varying assumptions about the relative magnitudes of the values of the joint benefits of an averting good.

Again, marginal willingness to pay cannot be determined from the household production technology alone because information on tastes enters the welfare expression. In this case the marginal averting expenditure does not incorporate the direct utility effect of environmental quality $((\partial U/\partial q)/\lambda)$ and therefore provides a lower bound but not an exact measure of the total value of the change in environmental quality.

8.5.2 Measuring Prices of Averting Actions

A second threat to valid welfare measurement is that many averting actions do not have easily determined prices. For example, averting behavior may require time, and the value of time varies among people and is difficult to measure, particularly for persons who are not employed. The price of medical care, a good often treated as the averting input in the health production approach, is difficult to determine because of health insurance coverage. Empirical researchers typically use some function of wage rates to value time and some measure of out-of-pocket expenses to measure the financial cost of medical care. It is less clear how to measure the price of substituting indoor for outdoor leisure activities.[28]

8.5.3 Identifying Partial Effects of Averting Behavior and Environmental Conditions

A third obstacle arises from the need to identify the partial effects of averting behavior and environmental quality to estimate welfare change using Eq. (8.17) (Gerking and Stanley 1986). The most direct application of an averting behavior model is to use the estimated parameters of the household production function in Eq. (8.7) to estimate marginal willingness to pay in either Eq. (8.17) or Eq. (8.11). For example, Joyce et al. (1989) estimated a production function for risk of infant mortality (h) that included as inputs prenatal medical care (I) and ambient concentrations of sulfur dioxide (q). They used estimated parameters to predict marginal willingness to pay for reduced sulfur dioxide concentrations based on an equation as in (8.17). Dickie (2005) estimated a health production function for children's days of school absence due to illness with medical care (I) as an averting input and used an equation as in (8.11) to value avoiding 1 day of illness-induced school loss.[29]

[28]Mansfield et al. (2006) addressed this problem using a stated preference survey to estimate the value of adjustments in time use.

[29]Another way to estimate the household production technology is to transform the production function to obtain the function giving the amount of averting input required to produce a given

Attempts to estimate a household production technology confront the challenges of identifying the partial effects of behavior and environmental conditions on the home-produced output. These challenges arise because behavior is self-selected and environmental quality is not randomly assigned.

To focus on one of these problems at a time, assume for now that environmental quality is assigned as if randomly and consider implications of endogenous choices of averting action. Return to the Cobb–Douglas production function $h = GI^{\gamma_I} q^{\gamma_q}$ and let $G = G_0 z^{\gamma_z}\varepsilon$, where $G_0 > 0$. Assume that h, I, q, z and ε are positive, continuous, random variables. In logarithmic form the production function is

$$\ln h = \gamma_0 + \gamma_I \ln I + \gamma_q \ln q + \gamma_z \ln z + \ln \varepsilon, \tag{8.37}$$

where $\gamma_0 = \ln G_0$. Assume that h represents a measure of health. The new variable, z, represents an observed input that affects health. In a study employing cross-sectional data on individuals, for example, z might represent a personal characteristic like age. In an actual empirical study, there could be many inputs other than environmental quality and averting behavior included in the specification of the household production function. The other new variable, ε, represents the influence of factors affecting health that are unobserved by the researcher, even though some of these factors are known by the individual. For example, ε includes the effects of unmeasured exogenous determinants of health, such as the individual's resistance to illness, as well as unmeasured endogenous inputs like health habits or unobserved averting inputs.

Given the specification of the household production function in Eq. (8.37), one would use estimated parameters and data to estimate marginal willingness to pay in Eq. (8.17) as $p\left(\hat{\gamma}_q/\hat{\gamma}_I\right)(I/q)$, where $\hat{\gamma}$ denotes an estimated parameter. Typically one would evaluate this expression at the sample means of I and q or compute the value for each individual in the sample and then average over the sample. This procedure requires identification of the partial effects of behavior (γ_I) and environmental quality (γ_q) on health.

To consider prospects for identifying the parameters of the household production function, assume that z is exogenous and maintain the assumption that environmental quality is assigned as if randomly. These assumptions mean that both $\ln z$ and $\ln q$ are uncorrelated with the regression disturbance $\ln \varepsilon$. The averting input, I, is endogenous by assumption, however, and therefore $\ln I$ is correlated with $\ln \varepsilon$. To understand this important point, consider how a utility-maximizing individual would

(Footnote 29 continued)

level of home-produced output at a specified level of environmental quality $I^D = I^D(h, q)$. Estimated parameters of this function are used to compute marginal willingness to pay based on Eq. (8.18). Gerking and Stanley (1986) implicitly solved the production relationship to express medical care use (I) as a function of measures of health (h) and ambient air pollution (q). They estimated the effect of ambient ozone on use of medical care and multiplied by the full price of medical care to estimate marginal willingness to pay to reduce ozone concentrations, as suggested by Eq. (8.18).

choose I. The individual would determine the optimal amount of I in light of all available information, including factors contained in ε. In this example, an individual with the quasi-linear utility function $U = x + uh^{\alpha}$ would choose averting behavior according to the Marshallian demand function $I^{*} = \left(G_0^{\alpha}\beta_I p^{-1} q^{\beta_q} z^{\beta_z} u\varepsilon^{\alpha}\right)^{1/(1-\beta_I)}$, where $\beta_z = \alpha\gamma_z$ and other symbols have been defined previously. Taking logs,

$$\ln I = \pi_{I0} + \pi_{Ip} \ln p + \pi_{Iq} \ln q + \pi_{Iz} \ln z + \ln v_I, \tag{8.38}$$

where $\pi_{Ip} = -1/\delta, \pi_{Iq} = \beta_q/\delta$, $\pi_{Iz} = \beta_z/\delta$, $\ln v_I = 1/\delta(\alpha \ln \varepsilon + \ln u)$, $\delta = 1 - \beta_I$, and π_{I0} represents the intercept.

The reduced-form behavioral function in Eq. (8.38) shows how the individual chooses averting action so that $\ln I$ is a linear function of $\ln \varepsilon$. It follows that in the household production Eq. (8.37), the regressor $\ln I$ and the disturbance $\ln \varepsilon$ are correlated (positively correlated provided that $\alpha/\delta > 0$). As is well known, the least-squares estimators of the parameters of Eq. (8.37) are biased and inconsistent when a regressor is correlated with the disturbance. Therefore, the estimator of willingness to pay is inconsistent.[30]

Possible solutions to the problem of identifying parameters of the household production function are (1) to find an estimator that is consistent in the presence of endogenous choices of averting behavior or (2) to estimate willingness to pay without use of the household production function. Consider first two potentially consistent estimators of parameters of the household production function in Eq. (8.37): instrumental variables and fixed effects.

A valid instrumental variable should be relevant (correlated with the endogenous averting input) and exogenous (uncorrelated with the disturbance in the household production function). It is quite difficult to find a valid instrument for averting action. Early research on averting behavior, in keeping with the econometric practice of the time, typically determined instruments based on exclusion restrictions imposed by the theoretical model. These restrictions specify variables that enter the reduced-form demand function for averting behavior (Eq. (8.38)) but do not enter the household production function (Eq. (8.37)).

To illustrate the application of an exclusion restriction that is based on the theoretical structure of the model, note that in Eq. (8.38), $\ln I$ depends on $\ln p$, but $\ln p$ is excluded from the household production function in Eq. (8.37). Provided the coefficient $\pi_{Ip} \neq 0$, the (log of) price of an averting input meets the condition of instrument relevance. Theoretically, p is a market price and thus is unrelated to observed or unobserved characteristics of the individual. Therefore, the price of an averting input may also meet the condition of instrument exogeneity.

Although the price of an averting input can be a valid instrument for the endogenous averting input in some circumstances, there are practical difficulties

[30]Similar problems hinder identification of parameters of the required defensive input function $I^D(h, q)$ and of the defensive expenditure function $D(h, p, q)$ because these functions are transformations of the household production function.

that can threaten this reasoning. First, the price must be observable, but the price of averting action often is not observed or is difficult to measure accurately. Second, the price must vary, but in the case of averting inputs purchased in markets, all individuals in a given location might face the same price. Third, variation in price must be uncorrelated with variation in the disturbance in the household production function, but this condition will often be violated.[31]

In general, exclusion restrictions imposed by theory can be helpful in identifying potential instrumental variables but are not by themselves sufficient to establish validity of the instruments.[32]

Another approach to constructing a valid instrument for averting action is to find variation in averting behavior that arises from essentially random sources. This quasi-experimental approach is difficult to apply because averting action is by assumption a choice made by the individual. A possible example might be a situation in which some people received information about a change in environmental quality (e.g., from a hazard warning), and others did not. If receipt of the information is independent of unobserved determinants of health and induces a behavioral response, then the information may represent a valid instrument.

A second potentially consistent estimator of the household production function arises when there are repeated observations on the same cross-sectional units. For example, suppose one intends to estimate a household production function using microdata on $i = 1, \ldots, N$ individuals, each observed for $t = 1, \ldots, T$ periods. Suppose that the disturbance in the household production function of Eq. (8.37), $\ln \varepsilon_{it}$, can be decomposed into an individual-specific, time-invariant component, $\ln \mu_i$, and a second component that varies over individuals and time, $\ln \omega_{it}$, so that $\ln \varepsilon_{it} = \ln \mu_i + \ln \omega_{it}$. The time-invariant component, $\ln \mu_i$, reflects the influence of individual-specific factors that do not vary over time, such as an individual's health endowment and persistent health habits. If these factors account fully for correlation between the disturbance and the averting input, then application of the fixed effects estimator for panel data (Greene 2012, Chapter 11) will provide consistent estimators of the effects of environmental quality and averting behavior in Eq. (8.37), under random assignment of environmental quality. If, however, the correlation between the disturbance and the averting input arises partly through the transitory error component, $\ln \omega_{it}$, which includes the effects of omitted, time-varying health

[31]Many applications use medical care as an averting input. The price of medical care depends on health insurance coverage, and insurance coverage is probably related to unobserved determinants of health (the disturbance in Eq. (8.37)). Also, many averting behaviors involve use of time and the full price of the averting input depends on the value of time. But the value of time is probably correlated with human capital and with unobserved determinants of health.

[32]Another exclusion restriction suggests income as an instrument for averting action. Were it not for the quasi-linear form of the utility function, individual income would appear in the Marshallian demand for the defensive input in Eq. (8.38), but income is not an input in the household production function. However, income is almost certainly correlated with unobserved determinants of health and thus would not satisfy instrument exogeneity.

inputs such as unobserved defensive behaviors, then fixed effects is not a complete solution.

For example, suppose that h represents a measure of acute exacerbations of asthma, q represents ambient air pollution, and I represents reductions in time spent outdoors. If the unobserved severity of chronic asthma differs between individuals but is fixed over time for any individual, then the fixed effects estimator avoids endogeneity problems arising from any correlation between severity and time spent outdoors. On the other hand, if unobserved use of asthma medication reduces acute exacerbations and varies over both individuals and time, correlation between medication use and time spent outdoors would cause inconsistency in the fixed effects estimator of the household production function. Using fixed effects estimation is likely to be a substantial benefit because it removes the effects of unobserved characteristics that are fixed over time, but it may not be a complete solution to the endogeneity problem because it does not remove the effects of unobserved factors that vary over both cross-sectional units and time.

In summary, identifying the partial effect of averting behavior on the home-produced output is difficult even under random assignment of environmental quality. Allowing for nonrandom assignment of environmental quality introduces an additional source of bias that further complicates identification of the partial effects of behavior and environmental conditions (see Sect. 8.5.4).

The difficulty of identifying the partial effects of behavior and environmental quality has prompted researchers to implement averting behavior methods without estimation of the household production function. As discussed in Sect. 8.3, there are two main ways to do this. One is to assume weak complementarity and input necessity and apply the Bockstael–McConnell approach to the estimated demand for an averting input. In the present example, this would involve estimation of Eq. (8.38). Under the assumed random assignment of environmental quality, parameters of this equation would be estimated consistently by least squares.

The second way to sidestep estimation of the household production function is to estimate the total effect of environmental quality on the home-produced output, $\mathrm{d}h/\mathrm{d}q$, and the effect of environmental quality on defensive behavior, $\partial I/\partial q$, to support estimation of marginal willingness to pay using Eq. (8.24). To develop an estimating equation for the total effect of environmental quality, substitute the demand for defensive input in Eq. (8.38) into the household production function in Eq. (8.37) to obtain the reduced-form damage function:

$$\ln h = \pi_{h0} + \pi_{hp} \ln p + \pi_{hq} \ln q + \pi_{hz} \ln z + \ln v_h, \qquad (8.39)$$

where $\pi_{hp} = -\gamma_I \delta$, $\pi_{hq} = \gamma_q/\delta$, $\pi_{hz} = \gamma_z/\delta$, $\ln v_h = 1/\delta(\ln \varepsilon + \gamma_I \ln u), \delta = 1 - \beta_I$, and π_{h0} represent the intercept.

The coefficient of the log of environmental quality in Eq. (8.39) represents the total effect of environmental quality on the log of the home-produced output: $\pi_{hq} = \gamma_q + \gamma_I \pi_{Iq}$. This is a linear version of Eqs. (8.1) and (8.23) showing that the total effect of environmental quality equals the partial effect with behavior held constant (γ_q) plus an indirect effect measured as the product of the effect of environmental quality on averting behavior (π_{Iq}) and the effect of averting behavior on health (γ_I).

Under the assumed random assignment of environmental quality, parameters of the reduced-form damage function would be estimated consistently by least squares. One estimates Eq. (8.39) to obtain an estimate of the total effect of environmental quality on health and then estimates Eq. (8.38) to determine the effect of environmental quality on defensive behavior. Applying Eq. (8.24) with the assumed functional forms, marginal willingness to pay for improved environmental quality is estimated as $((\partial U/\partial h)/\lambda)\hat{\pi}_{hq}(h/q) - p\hat{\pi}_{Iq}(I/q)$. The advantage here is that the reduced-form estimators of effects $(\hat{\pi})$ are consistent under assumed random assignment of environmental quality. A disadvantage, as discussed previously, is the need to apply a prior estimate of the marginal value of health $((\partial U/\partial h)/\lambda)$. A second disadvantage is that this approach provides no direct evidence that the averting input actually affects the home-produced output because no estimate of $\partial h/\partial I$ is produced.[33]

8.5.4 Identifying Total Effects of Environmental Conditions

The indirect approach of implementing the averting behavior method using the reduced-form damage and behavior functions avoids the difficulty of identifying partial effects in the household production function. But the approach still faces the challenge of identifying the total effect of environmental quality on health and the effect of environmental quality on behavior, in the face of nonrandom assignment of environmental conditions.

Pollution concentrations vary spatially, and people are not randomly sorted across space. Many other spatially differentiated factors that affect health (or other home-produced outcomes) may co-vary with pollution (Graff Zivin and Neidell 2013), including population density, racial composition, socioeconomic status, crime rates, health insurance coverage, and health choices such as diet and exercise. For example, people with lower resistance to illness might choose residential locations in less-polluted areas to avoid the health effects of air pollution. Such self-selection would impart a downward bias to the health effects of pollution estimated from cross-sectional variation. Alternatively, high pollution levels can be correlated with other unobserved threats to health, leading to an overstatement of the health effects of pollution.[34]

[33]See Neidell (2009) and Moretti and Neidell (2011) for approaches to work around the second shortcoming.

[34]Additionally, there can be substantial measurement error in the measures of pollution concentrations because of imperfect matching of individuals to pollution monitors and variations in the amounts of time spent indoors and outdoors. Measurement error would be expected to attenuate the estimated effect of pollution on health.

Three approaches to identifying the total effects of environmental conditions on health or on averting behavior will be illustrated in this subsection: (1) fixed effects, (2) instrumental variables for environmental conditions, and (3) regression discontinuity. The Deschesnes and Greenstone (2011) study of climate change and mortality illustrated the use of fixed effects to identify the total effect of environmental conditions on health. They estimated reduced-form equations for health (mortality) and for averting behavior (residential energy expenditures for cooling). They used many years of U.S. county-level data to estimate the total effect of temperature on mortality using their version of Eq. (8.39), shown in Eq. (8.40):

$$h_{cta} = TEMP'_{ct}\theta + PRECIP'_{ct}\delta + \alpha_{ca} + \gamma_{sta} + \varepsilon_{cta}. \tag{8.40}$$

In this equation, h_{cta} represents the mortality rate in county c, year t for age group a, and $TEMP_{ct}$ and $PRECIP_{ct}$, respectively, denote vectors of dummy-variable indicators for levels of temperature and precipitation for county c in year t, with associated vectors of coefficients θ and δ. Also, α_{ca} denotes a set of county-by-age-group fixed effects, γ_{sta} denotes a set of state-by-year-by-age-group fixed effects, and ε_{cta} is the disturbance. Inclusion of county-by-age-group fixed effects removes the influence of any county-specific, time-invariant determinants of mortality for a given age group. The state-by-year-by-age-group fixed effects control for any time-varying influences on mortality that vary between states but are common to all counties within a state for a given age group. Thus, the θ coefficients measuring the total effects of temperature on mortality are identified by within-county variations in temperature after removing linear effects of precipitation and of all factors common to a given age group in a given year in a given state.

Moretti and Neidell (2011) used fixed effects with an instrumental variables estimation to identify the effect of ambient ozone concentrations on health in the Los Angeles area. They specified a reduced-form damage function in which illness (daily respiratory-related emergency room visits) depends on current and lagged values of ozone concentrations, current and lagged values of other air pollutants and weather variables, controls for seasonality and other omitted temporal influences, and fixed effects for the age group and the postal zone of residence of emergency room patients. Thus, the total effect of ozone on illness is identified by short-run variation in ambient ozone within postal zones for each age group, with other pollutants, weather, and temporal factors held constant. The authors then used a measure of boat traffic at the Los Angeles port as an instrument for ambient ozone. They showed that boat traffic contributes to ozone pollution (instrument relevance) and argued that it is uncorrelated with other short-term influences on illness (instrument exogeneity).[35]

[35]On the basis of evidence that boat traffic is unrelated to participation in outdoor activities, they argued that the instrument isolates variation in ozone that is independent of averting behavior and therefore that the instrumental variables' estimator identifies the partial effect of ozone on health. Their instrumental variables' estimates of the effect of ozone are about four times larger than their least-squares estimates.

Neidell (2009) used a regression discontinuity design to identify the effect of information about environmental conditions (air quality alerts) on averting behavior (reduced attendance at outdoor facilities). In a regression discontinuity design, observations are assigned to treatment or control based whether a threshold value for an assignment variable is exceeded (Shadish et al. 2002; Angrist and Pischke 2009). Neidell's assignment variable was the day-ahead forecast of the ozone concentration, and the threshold value was 20 pphm; an air quality alert was issued if the forecast exceeded 20 pphm. If days with ozone forecasts just above 20 pphm (treatment observations) and just below 20 pphm (control observations) are otherwise identical, a discontinuous decline in outdoor activities at the threshold identifies the effect of the alert.

Neidell (2009) estimated two types of regression discontinuity models using the log of aggregate attendance at outdoor facilities as the dependent variable and controlling for realized air pollution and weather and for temporal factors affecting attendance. One model enters dummy indicators for the level of the ozone forecast. The difference between the coefficients of the dummy variables for ozone forecasts of 20 pphm and 19 pphm measures the effect of an alert. The other model replaces the dummy indicators for the forecast ozone level with (1) a dummy variable indicating whether an alert was issued and (2) the forecast ozone level. In this specification, the effect of an alert is identified by holding the ozone forecast constant. Neidell found a significant and sizeable decline in outdoor activities when an alert was issued.

8.6 Conducting an Averting Behavior Study

Table 8.3 presents a stylized sequence of steps for conducting an averting behavior study. Of course, every application is different, but the table lists many key decisions and actions.

8.6.1 Specify the Change to Be Valued

Begin by specifying the change to be valued in terms of the baseline and alternative conditions (Table 8.3, Step 1). Three separate types of values can be estimated using averting behavior methods. First, one can value a marginal change in an outcome or risk affected by the environment using a welfare measure as in Eq. (8.11) or Eq. (8.33). Second, one can estimate marginal willingness to pay for a change in environmental quality using an expression as in Eq. (8.17), Eq. (8.24), or Eq. (8.34). Third, one can estimate compensating or equivalent surplus for a nonmarginal change in environmental quality. When valuing a nonmarginal change, one might estimate the exact compensating surplus using the Bockstael–McConnell method (Eq. (8.28)) or estimate a bound on the surplus as described by

Table 8.3 Conducting an averting behavior study

Step 1	Specify the change to be valued
Step 2	Describe the setting and specify the conceptual framework
Step 3	Choose an averting behavior method, identification strategy, and empirical specification
Step 4	Collect data to implement the chosen approach
Step 5	Estimate the model and its welfare expression and assess validity

Bartik (1988; Eq. (8.30)). In this section, it is assumed that the goal is to value a marginal change in environmental quality.

8.6.2 Describe the Setting and Specify the Conceptual Framework

Decide how to model the effect of environmental quality on welfare (Table 8.3, Step 2). Practical considerations often will demand taking the simplest approach of assuming that effects of environmental quality operate only through the household production function for a single outcome valued by an individual. But one should consider how this assumption could influence the interpretation of results. For example, if improved environmental quality influences welfare through some means other than a single household production function, then Eq. (8.17), Eq. (8.24), and Eq. (8.34) will understate the overall marginal willingness to pay to improve environmental quality, as illustrated in Eq. (8.36).

Having determined the outcome through which environmental quality affects welfare, consider the inputs in the household production function for this outcome. The inputs that are central to the model are environmental quality and averting behavior. Often, several measures of environmental quality should be included as inputs to avoid omitted variables bias. For example, several ambient air pollutants could affect health and behavior and could co-vary with one another. If so, exclusion of some pollutants would cause omitted variables bias.

In terms of averting inputs, most prior applications have fallen into one of four categories: (1) reductions in outdoor activities to avoid exposure to air pollution, (2) use of medication or medical care to mitigate the adverse health effects of air pollution exposure, (3) purchase or use of safety equipment to reduce the risk of death or injury, and (4) use of bottled water, home filtration, boiling of tap water, hauling water from an alternative source, or substituting other beverages as averting behaviors to reduce exposure to contaminated supplies of drinking water. Although these examples encompass most prior applications, the specific valuation problem at hand can lead to other definitions of outcomes and averting behaviors, such as the use of sunscreen lotion to reduce the risk of skin cancer (Murdoch and Thayer

1990) or residential energy consumption to defend against the health effects of climate change (Deschenes and Greenstone 2011).

One should consider, at least conceptually, a variety of possible averting behaviors, including behaviors that reduce exposure to environmental hazards and behaviors that mitigate the effects of exposure. For example, prior research suggests that many people use more than one averting behavior to avoid incidents of water contamination (Harrington et al. 1989). The full defensive expenditure would not be captured if some averting behaviors were not measured (e.g., if the use of home filtration was counted, but purchases of bottled water were ignored). More subtly, omitted variables bias can arise from substitution between measured and unmeasured averting actions. For example, two ways to reduce the risk of skin cancer are to (1) reduce time spent outdoors in direct sunlight and (2) use sunscreen when outdoors. People using sunscreen tend to spend more time outdoors in direct sunlight than they otherwise would, thus offsetting some of the risk reduction offered by the sunscreen (Dickie and Gerking 1997, 2009). A researcher who estimates the effect of using sunscreen lotion on the risk of skin cancer without controlling for time spent outdoors would be expected to underestimate the risk reduction produced by using sunscreen.

The researcher should also consider the effects of other inputs, including other choice variables like health habits and exogenous conditions that may co-vary with pollution or averting behavior. For example, weather conditions affect health, influence behavior, and co-vary with air pollution. Thus, a study of the health effects of air pollution should control for weather conditions.

In light of the pervasive presence of joint production in the household production setting and the resulting threat to the validity of welfare measures based on averting behavior, one should consider how joint production can influence the interpretation of results. In most cases the influence of joint production on estimated willingness to pay will not be removed directly but will be assessed, as discussed in connection with Eq. (8.35).

An additional feature of the setting and conceptual framework is the issue of whether a single individual makes decisions about averting actions and experiences all the resulting costs and benefits. Many people live in multiperson households in which the decisions of one person can affect the welfare of another household member. A researcher might want to consider the household setting in which averting decisions are made to focus on a parent's choices of averting actions on behalf of a child or to incorporate averting inputs like water filtration systems or air purifiers that could be public goods within the household. Gerking and Dickie (2013) and Adamowicz et al. (2014) discussed relevant models of household decision-making.

8.6.3 Choose an Averting Behavior Method, Identification Strategy, and Empirical Specification

Based on the discussion of identification in Sect. 8.5, the most credible estimates of marginal willingness to pay for environmental quality probably would be obtained by focusing on the reduced-form damage function for the outcome considered and the reduced-form demand function for the behavioral input. Marginal willingness to pay would then be estimated using Eq. (8.24). This approach avoids the need to identify the partial effect of averting behavior on the home-produced output but faces the challenge of identifying total effects of environmental quality on the home-produced output and on averting behavior.

Specify the equations to be estimated by choosing the variables that enter the reduced-form damage function, the reduced-form behavioral function, and the functional forms of these equations. In theory, the explanatory variables and functional forms of these equations are determined by solving the first-order conditions of the utility maximization problem. Thus, the explanatory variables are all of the exogenous variables (those not chosen in the utility maximization problem) that appear in the utility function, household production function, and budget constraint. The functional forms likewise are determined by the forms of the utility function, household production function, and budget constraint.[36]

In practice, researchers rarely determine the specification of the damage function and averting behavior function by solving the first-order conditions explicitly. Instead, one uses the structural model to provide guidance regarding the co-variates to include, while also following good econometric practice for selection of regressors and functional form (Greene 2012, Chapters 5-6; Wooldridge 2010, Chapter 6). Often there is little prior information about appropriate functional forms for the damage and averting behavior functions. One would usually start with a simple functional form such as linear or logarithmic while allowing for some nonlinearity in the effect of environmental conditions on health outcomes and behavior. There are many ways to allow for nonlinear effects. One approach that allows flexibility in the estimated responses to changes in environmental conditions is to use dummy-variable coding for levels of environmental variables as shown in Eq. (8.40) (Deschenes and Greenstone 2011).

To estimate the reduced-form damage function and the reduced-form function governing the choice of averting behavior requires developing a strategy to identify the effects of environmental quality. The identification strategy should rely partly on use of a fixed effects estimator with panel data if possible, as discussed in Sect. 8.5.

[36]For example, the reduced-form behavioral and demand functions in Eqs. (8.38) and (8.39) are determined as solutions to the illustrative model of Sect. 8.3 in which the utility function is quasi-linear, the household production function is Cobb–Douglas, and the budget constraint is linear (see Eq. (8.16)).

When using microdata in which the cross-sectional units are individuals, use of individual-specific effects removes the influence of all factors that vary between individuals but are fixed over time (Greene 2012, Chapter 11). This will eliminate many potential sources of omitted variables bias, including unmeasured health habits, pre-existing stocks of human capital, chronic health status, resistance to illness, preferences for environmental quality, and fixed locational characteristics (if individuals do not relocate during the study).

When using more aggregated data, the cross-sectional units typically are spatial or jurisdictional units (e.g., counties or census tracts). As discussed in Sect. 8.5, many spatially differentiated factors that affect health or behavior may co-vary with environmental quality. Use of spatial or jurisdictional fixed effects removes the influence of factors that vary between locations but are fixed over time, eliminating an important source of omitted variables bias.

In both microdata and aggregate data, use of fixed time effects removes the influence of factors that vary temporally but are constant for all cross-sectional units. When both cross-sectional and temporal fixed effects are used, the scope for omitted variables bias is limited to uncontrolled factors that vary over cross-sectional units and over time and are correlated with health or averting behavior.

When a substantial threat of bias remains from co-variation of environmental conditions and factors that vary over cross-sectional units and over time (or when using cross-sectional data), another method of identifying the effects of environmental conditions should be considered. As discussed in Sect. 8.5, instrumental variables and regression discontinuity methods have been used for this purpose in averting behavior studies.

8.6.4 Collect Data

Issues of data collection for nonmarket valuation are discussed in detail in Chap. 3. One general consideration is the definition of the target population about which one wants to make inferences. One might use environmental monitoring data to identify the population affected by the environmental conditions or policy changes considered. For some environmental problems, the affected population is local or regional, such as in the case of localized incidents of water contamination (Harrington et al. 1989; Abdalla 1990; Dasgupta 2004) or regional dispersion of smoke from wildfires (Richardson et al. 2012). Other problems and policies affect a national or even international population, such as ozone depletion (Murdoch and Thayer 1990) or climate change (Deschenes and Greenstone 2011). In many cases the study population differs from the population about which one would like to make inferences, such as when a study of averting behavior and air pollution is conducted in one city, but one wishes to make inferences about welfare change in the national population. These situations call for careful consideration of external validity[37] (Shadish et al. 2002) and benefit transfer (Chap. 11).

In addition to the guidance on data collection for nonmarket valuation provided in Chap. 3, a few issues of particular importance for conducting an averting behavior study warrant further discussion here. First, use panel data where possible so that a fixed effects estimator can be used to aid identification of effects of environmental conditions.

Second, collect data on the home-produced outcome and on averting behavior. Usually the outcome is a measure of health. Acute morbidity, or short-term illnesses, can be measured by self-reports (Richardson et al. 2012); by encounters with the health care delivery system such as hospitalizations (Neidell 2009); or by outcomes such as absence from work or school due to illness (Dickie 2005). Chronic morbidity, or long-term illness, is typically measured as presence of physician-diagnosed conditions. Mortality, or death, is more objectively measured than presence of illness because death is a discrete event that is almost universally recorded.

Third, data on environmental conditions are needed. Obtaining appropriate environmental data can be a challenge. The temporal scale of the required data is dictated by the outcome. Analysis of chronic morbidity or mortality in adults would warrant measurement of long-term environmental exposures, but long-term exposure data are rarely available. In the case of acute illness, measures of daily variation in environmental conditions typically would be sufficient, perhaps with allowance for lagged effects of the environment on health and behavior. Environmental data then must be matched to exposures of people, which is usually done on the basis of residential location. For example, one matches individuals to the nearest air quality monitoring station or to some weighted average of pollution readings at nearby monitoring stations. Similarly, data on weather or other local attributes should be collected and matched.

Fourth, consider inclusion of additional controls for omitted influences that may be correlated with included variables. As discussed previously, the use of fixed effects will remove the influence of many factors. When using panel data on individuals, for example, indicators for personal characteristics such as gender, racial background, illness, measures of health habits that are constant over the time period considered, and other individual variables would not need to be measured because the influence of these factors is absorbed by the fixed effects. But data on other factors that vary over individuals and time, which could co-vary with included variables, should be collected if possible.

Fifth, decide whether to collect primary data or to use secondary data. Primary data collection has been used in a few health production or averting-input demand studies (Dickie and Gerking 1991a, b, 1997, 2009; Richardson et al. 2012) and in most defensive expenditure studies (Harrington et al. 1989; Abdalla 1990). Primary

[37] External validity concerns whether inferences can be generalized from the population and setting studied to other populations and settings.

data collection usually involves administration of a survey. This costly but flexible approach allows the researcher to collect data on the specific averting behaviors and outcomes of interest, on attitudes and perceptions, and on the prices of averting behaviors. But collecting a large, representative survey sample, particularly with repeated observations of the individuals over time, is quite expensive and time-consuming. Thus, primary data collection often forces the researcher to settle for small samples that may not be representative of the target population. In collecting primary data, general principles of sampling and survey design apply (see Chap. 3). Using secondary data avoids the time and expense of conducting a survey. In some cases, representative samples reflecting extensive quality control and high response rates may be available, sometimes with repeated observation of the same cross-sectional units over time. On the other hand, use of secondary data limits the averting behaviors and outcomes to those measured in the sources used.

Sixth, decide whether to use data at an individual or more aggregated level. Individual data best match the theory because the unit of analysis in the theoretical model is the individual. Microdata on individuals will more fully reflect the variation among persons in circumstances, behavioral choices, and outcomes experienced and will more easily support analysis of heterogeneity between demographic groups. Aggregate data, such as data at the zip code or county level, will suppress much of the natural variation but will often allow use of large sample sizes, multiple levels of fixed effects, and consistent measurement over a long time period.

8.6.5 Estimate Model, Interpret Results, and Assess Validity

The next step is to estimate the reduced-form damage function and the reduced-form equation describing the choice of averting behavior. Once the equations have been estimated and the price of the averting input measured, estimating marginal willingness to pay is a straightforward application of an expression as in Eq. (8.24). An important consideration in the empirical analysis is to assess the validity of the strategy to identify the effects of environmental conditions. A simple way to begin to assess validity is to test the sensitivity of results to variations in specification. If the model is well identified, results should be robust to reasonable changes in specification. Standard practices for assessing validity of instrumental variables (e.g., Murray 2006) or quasi-experimental methods (Shadish et al. 2002; Angrist and Pischke 2009) should be followed if these techniques are employed.

8.6.6 A Simple Empirical Example

This subsection presents an empirical example to illustrate the steps just described for implementing an averting behavior method. The illustration is kept simple to avoid obscuring the main points behind complexities of design or analysis.

The change to be valued is a marginal reduction in ambient ozone concentrations. The health outcome considered is a measure of acute illness, namely a count of the number of symptoms experienced over a 2-day period. Assume that individual utility depends positively on consumption of a market good and negatively on the number of symptoms; as a starting point, assume that averting behavior and ambient ozone affect welfare only through their influence on the number of symptoms. Later, the impact of joint production or nonhealth benefits of ozone reductions will be assessed qualitatively using Eqs. (8.35) and (8.36). Ignore intra-household interactions and assume that an individual maximizes utility subject to a household production function for the number of symptoms and budget and time constraints.[38]

The household production function specifies the number of symptoms as a function of measured and unmeasured inputs. Measured inputs include concentrations of air pollutants and averting behavior. Two averting behaviors are considered to illustrate different aspects of the method. The first is time spent outdoors during the 2-day period because most evidence on avoiding the effects of air pollution involves reductions in outdoor activities at high concentrations of ambient ozone. The second averting behavior is an indicator for whether medical care (a visit to physician's office, emergency care facility, or hospital) was obtained during the 2-day period because many applications have used medical care as the averting input, and this behavior has a measured price. Measured weather conditions are included as control variables. Individual fixed effects are used to control for measured or unmeasured inputs that vary between individuals but are fixed over time, such as chronic health status, propensity to experience symptoms, health habits, gender, and so on. Similarly, measured or unmeasured time-varying inputs common to all individuals are controlled using fixed effects for time. Unmeasured inputs that vary over both individuals and time are not controlled.[39]

The averting behavior method employed is to estimate a reduced-form damage function for the number of symptoms and reduced-form behavioral functions for time spent outdoors and use of medical care. The specification of these equations is guided by the model just outlined. Thus, co-variates are measures of pollution and weather, with fixed effects for individuals and time (dummy variables for month of year and day of week). The individual fixed effects absorb the influence of many additional exogenous variables, such as prices, wages, and incomes, which are assumed constant over the time period considered.

Identification of the total effects of ambient ozone in the equations to be estimated is based on short-term variation in ozone concentrations for a given individual on a given day of the week in a given month, holding constant weather conditions and other air pollutants. This identification strategy relies on fixed effects and the assumption that the variation in ozone just described is uncorrelated with

[38]A time constraint is needed because both averting behaviors considered momentarily involve use of time. See Dickie and Gerking (1991b) and Bresnahan et al. (1997).

[39]Bresnahan et al. (1997) discussed other averting inputs.

uncontrolled influences on symptoms, outdoor time, or medical care use that vary over both individuals and time.

Primary panel data on individuals were collected by survey during 1985-86 from residents of the Burbank and Glendora communities near Los Angeles. The data were fully documented in Dickie and Gerking (1991b) and Bresnahan et al. (1997). Briefly, data were collected in an initial in-person interview followed by subsequent telephone interviews that inquired about daily activities and self-reported symptoms experienced during the 2 days preceding the survey. The 928 total observations constitute an unbalanced panel with between two and five observations for each of 226 individuals. These data were matched to concurrent measures of ambient air pollution from the monitoring station nearest to each respondent's home and to concurrent measures of temperature and humidity. The four so-called criteria pollutants for which complete data were available were used in the analysis: ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2).

Illustrative estimates of the damage function for the number of symptoms and the behavioral function for outdoor time are presented in Table 8.4. A Poisson model is used for the count of symptoms, and a linear regression is used for time spent outdoors. Preferred specifications appear in columns labeled (4) and (5), which include individual-specific fixed effects, fixed effects for day of week and month of year,[40] and controls for weather conditions. Results in columns labeled (1), (2), and (3) are presented to illustrate empirically several previously discussed points about omitted variables bias.

Consider estimated coefficients of ambient ozone in Table 8.4. Introducing controls for temperature and humidity in column (2) markedly changes the magnitude and statistical significance of the estimated ozone coefficients in the damage function and the behavioral function, relative to column (1) which omits controls for weather.[41] Inclusion of individual-specific fixed effects in column (3) and temporal fixed effects in column (4) induces further substantial changes in magnitude and significance of estimated ozone coefficients in both equations. These results suggest that failure to control for weather conditions and for unmeasured time-varying or individual-specific factors may result in substantial omitted variables bias in the estimated effects of ozone on health outcomes and behavior.

The specification of the outdoor hours' equation in column (5) is included to illustrate in a simple way the possibility of nonlinear effects of pollution discussed in Sect. 8.6.3. Intuition and previous research with these data (Bresnahan et al. 1997) suggests that people might not reduce time spent outdoors as air pollution concentrations increase unless air quality already is poor. Column (5) allows for

---

[40]Day and month indicators are used for time fixed effects because sample members were contacted on different dates. Air pollution varies by day and season, as do unobserved determinants of behavior and symptoms.

[41]Ozone is not emitted directly by sources but is formed from interactions involving other pollutants, sunlight, and heat. Thus ozone is correlated with weather, and because weather also affects symptoms and outdoor activities, exclusion of weather variables would lead to omitted variables bias in estimated ozone coefficients.

nonlinear effects of pollution using a specification that is piecewise linear in ozone. The kink or "knot" in the regression line occurs at a concentration of 20 pphm. According to results in column (5), a 1-pphm increase in ambient ozone boosts time outdoors by 0.15 of an hour at concentrations below 20 pphm, but reduces time outdoors by 0.63 h at higher concentrations; both effects are significant at the 5% level.[42] The parallel specification of the damage function in column (5) confirms a positive impact of ozone on symptoms but shows no significant change in the ozone coefficient at high concentrations.

Results in columns (4) and (5) of Table 8.4 suggest that ambient ozone increases the number of symptoms experienced and, at high concentrations, reduces time spent outdoors. Attempting to use these estimates with Eq. (8.24) to infer willingness to pay to reduce ambient ozone would confront the problem of measuring a price for reduced outdoor time. Consequently, a second averting behavior is considered to illustrate estimation of marginal willingness to pay to reduce the ambient ozone concentration.

Table 8.5 presents illustrative estimates of the reduced-form damage function for symptoms and the reduced-form behavioral function for the use of medical care. The medical care equation is estimated by logit. Estimates in Table 8.5 are based on a subsample of 729 observations, including the 177 individuals who indicated in the initial interview that they experienced symptoms in smoggy conditions.[43] For simplicity, only the specification parallel to column (4) in Table 8.4 is presented (with fixed individual and time effects and controls for weather, but no nonlinearity in ozone). Estimates suggest that increases in ambient ozone increase the number of symptoms experienced and the propensity to obtain medical care. The estimated marginal effect of a 1-pphm increase in ozone is 0.0570 on the conditional mean of the number of symptoms and is 0.0225 on the probability of obtaining medical care.[44]

To infer marginal welfare change from the estimated marginal effects requires an estimate of the value of avoiding a symptom and the price of medical care (see Eq. (8.24)). These values are assumed to be \$97 for avoiding one symptom for 1

[42]At concentrations above 20 pphm, the estimated effect of a 1-pphm increase equals $0.153 - 0.7867 = 0.63$, with a standard error of 0.30. The estimated slope change occurring at 20 pphm should not be taken literally as indicating a discrete change in behavior because the choice of 20 pphm for the knot is arbitrary. Using a variety of piecewise linear and quadratic specifications, Bresnahan et al. (1997) found that outdoor time declines when ambient ozone rises above 12 to 14 pphm. The knot at 20 pphm is used for a comparison to results of Neidell (2009), who found that outdoor activities decline when the *forecast* ozone concentration exceeds 20 pphm, triggering an air quality alert.

[43]The coefficient of ozone in the medical care equation is not significant at 10% in the full sample.

[44]The conditional mean number of symptoms and the probability of obtaining medical care are nonlinear functions of the inner products of co-variates and coefficients in the Poisson and logit models; coefficients therefore do not measure marginal effects of co-variates on the expected number of symptoms or probability of medical care use. See Greene (2012, Chapters 17-18) regarding the computation of marginal effects in these models. In the present analysis, marginal effects are computed at the mean of all co-variates.

Table 8.4 Example of damage function (number of symptoms) and averting behavior function (hours spent outdoors), estimated coefficients (standard errors)

Specification	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Constant intercept?	Yes	Yes	No	No	No
Conditional on:					
Temperature, humidity?[a]	No	Yes	Yes	Yes	Yes
Individual effects?	No	No	Yes	Yes	Yes
Day, month effects?[b]	No	No	No	Yes	Yes
Co-variate[c]	Number of symptoms (Poisson)[d]				
O_3 (pphm)	−0.0057	0.0219***	0.0055	0.0216**	0.0250**
	(0.0045)	(0.0072)	(0.0077)	(0.0102)	(0.0111)
CO (ppm)	−0.0328***	−0.0430***	−0.0005	0.0067	0.0052
	(0.0123)	(0.0126)	(0.0231)	(0.0249)	(0.0249)
SO_2 (pphm)	0.0094	0.0095	0.0093	0.0074	0.0070
	(0.0060)	(0.0062)	(0.0076)	(0.0089)	(0.0089)
NO_2 (pphm)	0.0545***	0.0427***	0.0488***	0.0193	0.0190
	(0.0087)	(0.0088)	(0.0119)	(0.0146)	(0.0146)
$1(O_3 \geq 20) \times (O_3 - 20)$[e]					−0.0425
(pphm)					(0.0571)
Log-likelihood[f]	−2,631	−2,615	−1,119	−1,089	−1,089
Co-variate[c]	Hours spent outdoors (least squares)				
O3 (pphm)	0.2182***	0.0926	0.1368**	0.0952	0.1530**
	(0.0414)	(0.0671)	(0.0585)	(0.0718)	(0.0750)
CO (ppm)	0.2924***	0.3058***	0.2403	0.3230*	0.3032*
	(0.1094)	(0.1130)	(0.1587)	(0.1663)	(0.1659)
SO2 (pphm)	−0.1064**	−0.1261**	−0.1093**	−0.1097*	−0.1141**
	(0.0535)	(0.0551)	(0.0507)	(0.0581)	(0.0579)
NO2 (pphm)	−0.1746**	−0.1425*	−0.1561*	−0.0408	−0.0478
	(0.0787)	(0.0802)	(0.0803)	(0.0939)	(0.0936)
$1(O_3 \geq 20) \times (O_3 - 20)$ [e]					−0.7867**
(pphm)					(0.3148)
R-squared	0.031	0.037	0.559	0.597	0.601

[a]Temperature is measured as the average of the daily high temperature over the 2-day period. Humidity is measured as the average of the daily low humidity, over the 2-day period
[b]Dummy variables for day of week and month of year of each contact with individual
[c]Pollutants are measured as the average of the daily maximum one-hour concentration over the 2-day survey period. pphm = parts per hundred million; ppm = parts per million
[d]Poisson models with individual effects are estimated by maximizing the conditional log-likelihood
[e]Variable equals 0 if $O_3 < 20$ pphm and otherwise equals $(O_3 - 20)$
[f]Log-likelihood values are not comparable for Poisson models with and without individual effects
Note Asterisks denote statistically significant coefficients at the *0.10, **0.05, or ***0.01 level in a two-tail test

Table 8.5 Example of damage function (number of symptoms) and averting behavior function (medical care consumption), estimated coefficients (standard errors in parentheses)[a]

Co-variate	Number of symptoms (Poisson)	Medical care (Logit)
O_3 (pphm)	0.0236**	0.1709*
	(0.0107)	(0.0962)
CO (ppm)	0.0020	−0.3226*
	(0.0261)	(0.1913)
SO_2 (pphm)	0.0100	−0.0317
	(0.0096)	(0.0741)
NO_2 (pphm)	0.0170	0.0041**
	(0.0156)	(0.0923)
Log-likelihood	−925	−46

[a]Estimates are conditional on individual-specific fixed effects, day and month effects, and temperature and humidity and are based on maximizing the conditional log-likelihood for the Poisson or logit fixed effects models for panel data. Estimates based on subsample of 729 observations, including 177 individuals who reported a tendency to experience symptoms in smoggy conditions in the initial interview

Note Asterisks denote statistically significant coefficients at the *0.10 or **0.05 level in a two-tail test

day and \$77 for the full (time-inclusive) price of a medical visit.[45] Using the estimated marginal effects and monetary values, illustrative estimates of the value of reducing the maximum one-hour daily concentration of ambient ozone by 1 pphm for 1 day can be determined by applying Eq. (8.24). The damage function estimate of the value of reducing ozone by 1 pphm for 1 day is $\$97 \times 0.057 = \$5.53$ per person. The estimated savings in averting expenditure from the same reduction in ozone is $\$77 \times 0.0225 = \$1.73$ per person, an amount 31% as large as the damage function estimate of value. Thus, estimated marginal willingness to pay to reduce ambient ozone is $\$5.53 + \$1.73 = \$7.26$ per pphm per person per day. The damage function estimate of value understates the willingness to pay estimate by 24%.

Although estimates presented are illustrative, it is worth completing the example by considering some of the main threats to validity. Key threats to external validity arise from the age of the data and the sample of individuals that is unrepresentative of the U.S. population (Dickie and Gerking 1991b). Concerning the validity of the willingness-to-pay measure, it would represent at best a partial measure of the overall welfare gain from reducing ozone, because it accounts only for the effect of ozone on acute symptoms while ignoring other health effects and any nonhealth impacts of ozone on welfare (see Eq. (8.36)). On the other hand, to the extent that

[45]The value of avoiding a symptom is computed as the average of values obtained by Dickie and Messman (2004) and Richardson et al. (2012) after inflating these values to 2014 dollars. The full price of medical care, computed as the out-of-pocket expense of a visit to one's regular physician plus the product of the wage and the time usually required for the doctor visit, was measured in the initial interview, averaged over the sample, and inflated to 2014 dollars. See Dickie and Gerking (1991b).

medical care provides joint benefits apart from mitigation of the health effects of ozone exposure, the willingness-to-pay estimate would overstate the true welfare improvement from reduced ozone arising from its impact on symptoms (see Eq. (8.35)). Furthermore, threats to internal validity arise from the econometric analysis presented in Tables 8.4 and 8.5. The identification strategy fails if there is any unmeasured determinant of symptoms, outdoor hours, or medical care that varies over individuals and time and is correlated with ozone;[46] if the functional forms are misspecified; or if there is measurement error arising, for example, from the assignment of pollution monitoring stations to individuals. Finally, the method of estimating reduced-form damage behavioral functions implemented here provides no direct evidence whether medical care alleviates symptoms of ozone exposure, but assumption is the basis of the welfare estimate.[47]

8.7 Concluding Comments

Averting behavior methods potentially offer ways to improve measurement of the physical effects of environmental conditions and to estimate the benefits of environmental improvement. Validity of measures of physical effects and benefits derived from averting behavior methods depends on how the challenges of joint production, unknown prices of averting actions, and identification of the effects of pollution and averting behavior are handled.

References

Abdalla, C. W. (1990). Measuring economic losses from ground water contamination: An investigation of household avoidance cost. Journal of the American Water Resources Association, 26, 451-463.

Adamowicz, W., Dickie, M., Gerking, S., Veronesi, M. & Zinner, D. (2014). Household decision-making and valuation of environmental health risks to parents and their children. Journal of the Association of Environmental and Resource Economists, 1, 481-519.

Agee, M. D. & Crocker, T. D. (1996). Parental altruism and child lead exposure: Inferences from the demand for chelation therapy. The Journal of Human Resources, 31, 677-691.

Akerman, J., Johnson, F. R. & Bergman, L. (1991). Paying for safety: Voluntary reduction of residential radon risks. Land Economics, 67, 435-446.

[46]An initial list of suspected sources of omitted variables bias might include lagged values of pollution and weather variables, additional measures of weather conditions, measures of allergens, and uncontrolled adjustments in behavior, such as changing the time of day of outdoor activities to avoid exposure to the peak daily concentrations of air pollution.

[47]Readers interested in seeing ways to address some threats to validity of the nature just outlined should consult studies of the health effects of ozone by Neidell (2009) and Moretti and Neidell (2011).

Angrist, J. D. & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.

Bartik, T. J. (1988). Evaluating the benefits of nonmarginal reductions in pollution using information on defensive expenditures. Journal of Environmental Economics and Management, 15, 111-127.

Baumol, W. J. & Oates, W. E. (1988). The theory of environmental policy (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.

Berger, M. C., Blomquist, G. C., Kenkel, D. & Tolley, G. S. (1987). Valuing changes in health risk: A comparison of alternative measures. Southern Economic Journal, 53, 967-984.

Blomquist, G. (1979). Value of life saving: Implications of consumption activity. Journal of Political Economy, 87, 540-558.

Blomquist, G. C. (2004). Self-protection and averting behavior, values of statistical lives, and benefit cost analysis of environmental policy. Review of Economics of the Household, 2, 89-110.

Bockstael, N. E. & McConnell, K. E. (1983). Welfare measurement in the household production framework. American Economic Review, 73, 806-814.

Bresnahan, B. W. & Dickie, M. (1995). Averting behavior and policy evaluation. Journal of Environmental Economics and Management, 29, 378-392.

Bresnahan, B. W., Dickie, M. & Gerking, S. (1997) Averting behavior and urban air pollution. Land Economics, 73, 340-357.

Chay, K. Y. & Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. Quarterly Journal of Economics, 118, 1121-1167.

Coase, R. H. (1960). The problem of social cost. Journal of Law and Economics, 3, 1-44.

Courant, P. N. & Porter, R. C. (1981). Averting expenditure and the cost of pollution. Journal of Environmental Economics and Management, 8, 321-329.

Cropper, M. L. (1981). Measuring the benefits from reduced morbidity. American Economic Review, 71, 235-240.

Dardis, R. (1980). The value of a life: New evidence from the marketplace. American Economic Review, 70, 1077-1082.

Dasgupta, P. (2004). Valuing the health damages from water pollution in urban Delhi, India: A health production function approach. Environment and Development Economics, 9, 83-106.

Deschenes, O. & Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. American Economic Journal: Applied Economics, 3 (4), 152-185.

Dickie, M. (2003). Defensive behavior and damage cost methods. In P. A. Champ, K. J. Boyle & T. C. Brown (Eds.), A primer on nonmarket valuation (pp. 395-444). Dordecht, The Netherlands: Kluwer Academic Publishers.

Dickie, M. (2005). Parental behavior and the value of children's health: A health production approach. Southern Economic Journal, 71, 855-872.

Dickie, M. & Gerking, S. (1991a). Valuing reduced morbidity: A household production approach. Southern Economic Journal, 57, 690-702.

Dickie, M. & Gerking, S. (1991b). Willingness to pay for ozone control: Inferences from the demand for medical care. Journal of Environmental Economics and Management, 21, 1-16.

Dickie, M. & Gerking, S. (1997). Genetic risk factors and offsetting behavior: The case of skin cancer. Journal of Risk and Uncertainty, 15, 81-97.

Dickie, M. & Gerking, S. (2009). Family behavior: Implications for health benefits transfer from adults to children. Environmental and Resource Economics, 43, 31-43.

Dickie, M. & Messman, V. L. (2004). Parental altruism and the value of avoiding acute illness: Are kids worth more than parents? Journal of Environmental Economics and Management, 48, 1146-1174.

Doyle, J. K., Mclelland, G. H., Schulze, W. D., Elliott, S. R. & Russell, G. W. (1991). Protective responses to household risk: A case study of radon mitigation. Risk Analysis, 11, 121-134.

Freeman, A. M., III. (1979). The benefits of air and water pollution control: A review and synthesis of recent estimates. Washington, DC: U.S. Environmental Protection Agency.

Garcia, P., Dixon, B. L., Mjelde, J. W. & Adams, R. M. (1986). Measuring the benefits of environmental change using a duality approach: The case of ozone and Illinois cash grain farms. Journal of Environmental Economics and Management, 13, 69-80.

Gerking, S. & Dickie, M. (2013). Valuing reductions in environmental risks to children's health. Annual Review of Resource Economics, 5, 245-260.

Gerking, S. & Stanley, L. R. (1986). An economic analysis of air pollution and health: The case of St. Louis. The Review of Economics and Statistics, 68, 115-121.

Graff Zivin, J. & Neidell, M. (2009). Days of haze: Environmental information disclosure and intertemporal avoidance behavior. Journal of Environmental Economics and Management, 58, 119-128.

Graff Zivin, J. & Neidell, M. (2013). Environment, health and human capital. Journal of Economic Literature, 51, 689-730.

Greene, W. H. (2012). Econometric analysis (7th ed.). Boston: Prentice Hall.

Harrington, W., Krupnick, A. J. & Spofford, W. O. (1989). The economic losses of a waterborne disease outbreak. Journal of Urban Economics, 25, 116-137.

Harrington, W. & Morgenstern, R. D. (2004). Evaluating regulatory impact analyses. Discussion paper 04-04. Washington, DC: Resources for the Future.

Harrington, W. & Portney P. R. (1987). Valuing the benefits of health and safety regulation. Journal of Urban Economics, 22, 101-112.

Hori, H. (1975). Revealed preference for public goods. American Economic Review, 65, 978-991.

Jakus, P. M. (1994). Averting behavior in the presence of public spillovers: Household control of nuisance pests. Land Economics, 70, 273-285.

Jones-Lee, M. W., Hammerton, M. & Philips, P. R. (1985). The value of safety: Results of a national sample survey. The Economic Journal, 95, 49-72.

Joyce, T. J., Grossman, M. & Goldman, F. (1989). An assessment of the benefits of air pollution control: The case of infant health. Journal of Urban Economics, 25, 32-51.

Lleras-Muney, A. (2010). The needs of the army: Using compulsory relocation in the military to estimate the effect of air pollutants on children's health. Journal of Human Resources, 45, 549-590.

Mansfield, C., Johnson, F. R., & Van Houtven, G. (2006). The missing piece: Valuing averting behavior for children's ozone exposures. Resource and Energy Economics, 28, 215-228.

McKitrick, R. & Collinge, R. A. (2002). The existence and uniqueness of optimal pollution policy in the presence of victim defense measures. Journal of Environmental Economics and Management, 44, 106-122.

Mishan, E. J. (1971). Evaluation of life and limb: A theoretical approach. Journal of Political Economy, 79, 687-705.

Moretti, E. & Neidell, M. (2011). Pollution, health, and avoidance behavior: Evidence from the ports of Los Angeles. Journal of Human Resources, 46, 154-175.

Murdoch, J. C. & Thayer, M. A. (1990). The benefits of reducing the incidence of nonmelanoma skin cancers: A defensive expenditures approach. Journal of Environmental Economics and Management, 18, 107-119.

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. Journal of Economic Perspectives, 20 (4), 111-132.

Mushkin, S. J. & Collings, F. d'A. (1959). Economic costs of disease and injury. Public Health Reports, 74, 795-809.

Neidell, M. (2004). Air pollution, health and socio-economic status: The effect of outdoor air quality on childhood asthma. Journal of Health Economics, 23, 1209-1236.

Neidell, M. (2009). Information, avoidance behavior and health: The effects of ozone on asthma hospitalizations. Journal of Human Resources, 44, 450-478.

Oates, W. E. (1983). The regulation of externalities: Efficient behavior by polluters and victims. Public Finance, 38, 362-375.

OECD. (2011). Valuing mortality risk reductions in regulatory analysis of environmental, health and transport policies: Policy implications. Paris: OECD.

Pollack, R. A. & Wachter, M. (1975). The relevance of the household production function and its implications for the allocation of time. Journal of Political Economy, 83, 255-278.

Pope, C. A. (1989). Respiratory disease associated with community air pollution and a steel mill, Utah Valley. American Journal of Public Health, 79, 623-628.

Quiggin, J. (1992). Risk, self-protection and ex ante economic value – some positive results. Journal of Environmental Economics and Management, 23, 40-53.

Rice, D. P. (1966). Estimating the cost of illness. Health Economics Series, No. 6. Washington, DC: U.S. Public Health Service.

Richardson, L. A., Champ, P. A., & Loomis, J. B. (2012). The hidden cost of wildfires: Economic valuation of health effects of wildfire smoke exposure in Southern California. Journal of Forest Economics, 18, 14-35.

Schelling, T. C. (1968). The life you save may be your own. In S. B. Chase (Ed.), Problems in public expenditure analysis: Papers presented at a conference of experts held Sept. 15-16, 1966 (pp. 127-162). Washington, DC: Brookings Institution.

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Belmont, CA: Wadsworth Cengage Learning.

Shibata, H. & Winrich, J. S. (1983). Control of pollution when the offended defend themselves. Economica, 50, 425-437.

Shimshack, J, P., Ward, M. B. & Beatty, T. K. M. (2007). Mercury advisories: Information, education, and fish consumption." Journal of Environmental Economics and Management, 53, 158-179.

Shogren, J. F. & Crocker, T. D. (1991). Risk, self-protection, and ex ante economic value. Journal of Environmental Economics and Management, 20, 1-15.

Smith, V. K. (1991). Household production functions and environmental benefit estimation. In J. B. Braden & C. D. Kolstad (Eds.), Measuring the demand for environmental quality (pp. 41-75). Bingley, United Kingdom: Emerald Group.

Smith, V. K., Desvousges, W. H. & Payne, J. W. (1995). Do risk information programs promote mitigating behavior? Journal of Risk and Uncertainty, 10, 203-221.

Thaler, R. & Rosen, S. (1976). The value of saving a life: Evidence from the labor market. In N. E. Terleckyj (Ed.), Household Production and Consumption (pp. 265-302). Cambridge, MA: National Bureau of Economic Research.

U.S. EPA (Environmental Protection Agency). (1997). The benefits and costs of the Clean Air Act, 1970 to 1990 – Study design and summary of results. Prepared by the Office of Administration and Resources Management, Office of Policy, Planning, and Evaluation, for the U.S. Congress. Washington, DC: U.S. EPA.

U.S. EPA (Environmental Protection Agency). (2012). Regulatory impact analysis for the proposed revisions to the national ambient air quality standards for particulate matter. Office of Air Quality Planning and Standards. EPA-452/R-12-003. Research Triangle Park, NC: U.S. EPA.

Watson, W. & Jaksch, J. (1982). Air pollution: Household soiling and consumer welfare losses. Journal of Environmental Economics and Management, 9, 248-262.

Weisbrod, B. A. (1971). Costs and benefits of medical research: A case study of poliomyelitis. Journal of Political Economy, 79, 527-544.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). Cambridge, MA: MIT Press.

Zeckhauser, R. & Fisher, A. (1976). Averting behavior and external diseconomies. Discussion paper 41D, J. F. Kennedy School of Government, Harvard University.

Chapter 9
Substitution Methods

Thomas C. Brown

Abstract This chapter covers two nonmarket valuation methods founded on the concept of substitution—the replacement cost method and equivalency analysis. Both methods are widely applied but frequently misunderstood. The replacement cost method is commonly used to value ecosystem services, and equivalency analysis is typically employed to determine required compensation for natural resource losses. The chapter describes and contrasts the methods, clarifies when (and when not) to apply them, illustrates the steps of a successful application, and provides examples of past applications. The reader will gain an appreciation of when the methods can provide a valid measure of economic value, and alternatively when the methods more usefully provide only a measure of cost. Following the guidance provided here can enhance the credibility of empirical applications of substitution methods.

Keywords Replacement cost method · Alternative cost method · Resource equivalency analysis · Habitat equivalency analysis · Substitutes · Valuation · Resource losses · Resource compensation · Public project · Natural resource damage assessment

Substitution methods offer markedly different valuation opportunities than those of earlier chapters of this text. Those chapters describe methods that focus on the behavior (or survey responses) of consumers, and thus on demand-side data. In contrast, the methods described in this chapter use data from the supply side about the characteristics and production costs of the goods and services desired by consumers.

T.C. Brown (✉)
U.S. Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA
e-mail: thomas.brown@colostate.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_9

Table 9.1 Uses of the substitution methods

Method	Aims to determine the
Replacement cost method	Value of a service provided by public investment to improve an existing situation (originally the alternative cost method)
	Value of an existing service that can be protected from loss by public effort
Equivalency analysis	Cost of compensating the public for losses following resource damage
	Cost of compensating the public for future resource losses due to development

Substitution methods rely on a deceptively simple condition—the availability of an alternative way of providing the same service as the environmental or other publicly provided service of interest.[1] Under certain conditions, the cost of providing the service in the alternate way can be taken as a measure of the value of the service of interest. And when the conditions for estimating an economic value are not met, substitution approaches still serve a useful role, in providing least-cost replacements for lost resources.

Two substitution methods are described here: the replacement cost method (RCM) and resource equivalency analysis.[2] The RCM dates back to the early decades of the twentieth century.[3] Originally known as the alternative cost method, the replacement cost approach was conceived as a way to value services provided by proposed public projects (Table 9.1), where the services would otherwise be provided by the private sector, such as valuing water-based shipping of goods made possible by controlling stream flow levels, where the shipping would otherwise be

[1]From here on, "service" is used to indicate both goods and services, and "environmental services" include ecosystem services.

[2]Before delving into the substitution approaches, it should be noted that other supply-side nonmarket valuation methods exist, most importantly the production factor method. The production factor method computes the value of a nonmarket input in production of items offered for sale. If changes in an environmental condition affect the output levels or costs of firms, producers' responses to such changes ultimately affect producers' and potentially also consumers' surpluses. Good introductions to the production factor method include Freeman (2003, Chapter 9), Point (1994), and Young (2005, Chapter 3). Because that method is conceptually more straightforward than substitution methods and is relatively uncontroversial, this chapter is devoted to the substitution methods.

[3]In the United States, the alternative cost method was in use by the early 1930s and increased in prominence with passage of the Flood Control Act of 1936, which stated that the feasibility of a proposed project is assured only when "the benefits, to whomsoever they may accrue, are in excess of the estimated costs." This statement essentially provided legislative support for the use of benefit-cost analysis in evaluation of public flood control projects, support that was later extended to other water projects (Griffin 2012). As U.S. government agencies began to implement the Flood Control Act and subsequent guidance in 1940s and 1950s, they searched for ways to measure the benefits of a variety of water resource projects, such as constructing flood control levees, dredging to improve navigation, and damming rivers to provide for water storage. Since then, the use of benefit-cost analysis has become common in the United States, Europe, and elsewhere, is used to evaluate many kinds of public projects, and takes advantage of the full range of valuation methods.

provided by private trucking or railroad firms. Later applications extended the method beyond the requirement that the alternative be provided by the private sector while still focusing on valuing proposed public projects, such as valuing the decrease in coastal flooding that would result from planting a mangrove forest, where the flooding would otherwise be controlled via sea walls and other constructions. In 1980s, as concern increased about the loss of environmental resources, the method was adapted to value possible environmental losses, and the "replacement cost" rubric was used (Table 9.1). This newer application of the replacement cost concept has been used, for example, to value the water quality maintenance made possible by protecting a wetland, which would otherwise be provided by a downstream drinking water treatment plant.

Somewhat different, but related, is equivalency analysis, which arose in 1990s in response to legislation requiring replacement of lost environmental services. Equivalency analysis is used to determine the required compensation either following damage to protected natural resources or preceding a planned loss of natural resource services, such as can happen when an area is developed (Table 9.1).

The RCM can produce valid estimates of economic value if certain conditions, explained further on, are met. As will become clear, these conditions are more likely to be met when the goal is to value a public infrastructure investment than when it is to value an existing resource that could be lost or diminished. In contrast, equivalency analysis is not intended to produce an economic estimate of value, but rather the cost of providing an equivalent resource service flow. Both methods provide estimates of the cost of providing, protecting, or compensating for lost environmental or publicly provided services.

9.1 The Replacement Cost Method

In a book on the use of benefit-cost analysis in water resource planning, Eckstein (1958) described an alternative cost concept that was being used to evaluate public water projects. He explained that when benefits cannot be estimated by observation of (or imputation from) market prices, certain conditions may exist where "benefit is taken to be equal to alternative cost, the cost of providing comparable output by the cheapest alternative means" (p. 52). The method was viable only where the objective of the project under study would be met with or without the project, meaning that if the proposed project were not undertaken, the same service would be provided in an alternate way. He assumed that the alternative would be provided by the private sector, such as when private railroads haul goods that would be transported by barge if a proposed publicly financed dredging operation made river navigation possible.

Use of the replacement cost approach to measure the benefit of a proposed public project, Eckstein explained, relies on two key principles: (1) consumers would not pay more for a service than the cost of a perfect substitute, and (2) if provision of the service via a public project would preclude the otherwise certain expenditure of

Table 9.2 Conditions for equating the benefit of a proposed action (Option 1) with the cost of an alternative (Option 2)

	Condition
Condition 1	Option 2 provides the same benefit as Option 1
Condition 2	Option 2 is the next least-cost alternative for providing the service (i.e., $c_1 \leq c_2$ and c_2 is the lowest alternative per-unit cost)
Condition 3	In the absence of option 1, the substitute would be demanded, if available, up to some specified quantity, at cost c_2 (i.e., $b_2 \geq c_2$)

the cost of the alternative, it would release those funds for use elsewhere in the economy and increase national income by approximately an equivalent amount. These principles are fundamental to acceptance of the RCM as a viable approach for estimating the economic value of certain pubic construction or protection efforts. Equally important, however, are the three essential conditions for applying the method, explained in the next section.

9.1.1 Necessary Conditions for Applying the Replacement Cost Method

The RCM equates the value of the service that a proposed project would provide with the cost of providing that service by other means. Acceptance of a cost as an estimate of the benefit of a project relies on meeting the three conditions listed in Table 9.2.[4] The conditions assume there are two options for providing the service—the public project of interest (Option 1) and the alternative (Option 2). Let b_1 and b_2 be the per-unit benefits of the two options and c_1 and c_2 be their per-unit costs. The quantity b_1 is the unknown value of interest. Condition 1 is that both options provide the same service—that they are perfect (or very close) substitutes. Condition 2 is that $c_1 \leq c_2$ but that c_2 is the least-cost alternative to c_1. Condition 3 requires that $b_2 \geq c_2$, indicating that Option 2 would be demanded if it were available at c_2 and Option 1 were not available. The logic behind the method is then as follows: If Condition 3 is met, we know that people would demand the service from Option 2 if Option 1 were not available and Option 2 could provide the service at a cost of c_2 (thus, $b_2 \geq c_2$). If Condition 1 is met, we know that people would be equally satisfied with Option 1 or Option 2, so it must also be that $b_1 \geq c_2$; however, because people would not pay more for something than the cost of a perfect substitute, in the presence of Option 2, $b_1 \leq c_2$. Thus, it must be that $b_1 = c_2$.[5] Finally, if Condition 2 is met, we know that c_2, and therefore b_1, is not being overstated.

[4]These three conditions were described by Eckstein (1958) and later summarized by Shabman and Batie (1978).

[5]It is sometimes argued that the replacement cost is a lower bound, not an upper bound, on the value of the project or resource at issue. For example, an EPA report (2009, p. 52) states that when

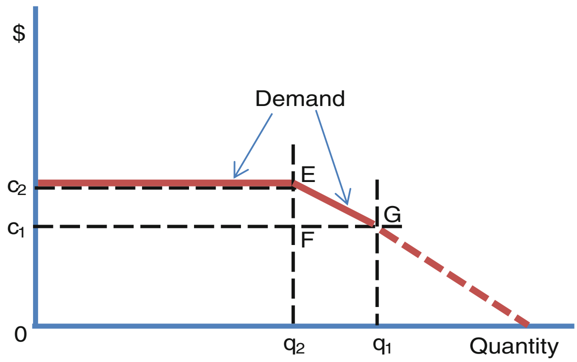


Fig. 9.1 Annual demand and supply given availability of a more expensive perfect substitute

To further illustrate the RCM, consider Fig. 9.1, representing conditions in a given year. Assume that a private alternative is able to supply the service at a constant price of c_2 and that the public project would be able to supply the same service at a constant cost of c_1. In the absence of the public project, q_2 of the service would be consumed. If the public project supplied any quantity less than or equal to q_2, the cost savings per unit would be c_2–c_1. For example, assume that the public project supplied exactly q_2 of the service. Demand for the service that the public project would provide is perfectly elastic at price c_2 up to output q_2 because if the public project were not undertaken, the service would be provided by the private alternative at a price of c_2. The annual gross benefit of the public project equals $b_1 \times q_2$, which is equivalent to $c_2 \times q_2$ and represented by the area $0c_2\mathrm{E}q_2$. The net benefit, equal to the cost savings, is shown as area $c_1c_2\mathrm{EF}$. Under these conditions, we have determined the maximum willingness to pay for the proposed project without having to estimate the underlying demand curve for the service because the effective demand curve is bounded by the cost of the perfect substitute.[6] Of course, in the absence of a perfect substitute (i.e., if the service provided by Option 2 differed in some ways from the service provided by the public project), the demand curve for the service would not necessarily be captured by the price of the

(Footnote 5 continued)

Conditions 1 and 3 are met, "It is valid to use the cost of providing the equivalent services via the alternative as a lower-bound estimate of the economic value of the ecosystem service." This argument essentially says that if Condition 3 is met, we know that society is willing to pay at least the replacement cost to have the service provided, and thus the original ecosystem service or publicly provided resource must also be worth at least that much. While this is surely true, the argument ignores the limitation that the available perfect substitute places on maximum willingness to pay. If one accepts that willingness to pay is limited by the cost of an available perfect substitute, the gross benefit of the publicly provided resource cannot exceed the replacement cost.

[6]A depiction of aggregate demand for the original service—that is, the service in the absence of a perfect substitute—would naturally be shown by a downward sloping demand curve. The horizontal demand curve in Fig. 9.1 depicts the situation faced by a new, or alternative, supplier of the identical service.

substitute. For example, if the service provided by Option 2 lacked a desirable attribute available with Option 1, the demand curve of the service provided by Option 1 cannot be bounded by c_2 and the gross benefit of Option 1 could exceed $c_2 \times q_2$.

Now allow for the possibility that the public project could produce more than q_2 of the service at a cost of c_1. In this case, and assuming again that Option 2 provides a perfect substitute, an additional segment of the demand curve must be estimated in order to compute the quantity demanded and the additional benefit. If q_1 of the service could be produced by the public project at cost c_1, the segment EG (Fig. 9.1) would need to be estimated, which would allow for adding the additional gross benefit indicated by area q_2EGq_1 and the additional net benefit FEG. To summarize, for output up to q_2, the benefit from the public project is limited and measured by the private cost, c_2, because beneficiaries would be willing to pay only 0c_2Eq_2 for public provision of the output at q_2. Adding the additional output, the gross benefit with the public project is 0c$_2$EGq_1, and the net benefit is c_1c_2EG.

Conditions 1 and 3 were specified above in terms of per-unit costs and benefits, as was the conclusion that if the conditions are met, then $b_1 = c_2$. Unit costs, and therefore benefits, would be observed if the alternative were provided by the private sector at a constant unit cost (which is presumed to reflect the full social cost of providing the service). In the case where the cost of the alternative is not yet observed or is expected to change over time and would need to be computed based on estimates of future fixed and variable costs over the planning horizon, the present value of the full cost over the planning horizon would be compiled as follows:

$$PV(C_2) = \sum_{t=0}^{T} C_{2,t}\rho^{-t}, \tag{9.1}$$

where $C_{2,t}$ is the annual total cost of Option 2 in year t, the planning horizon is T years, and $\rho = 1 + r$, where r is the discount rate. The present value of the cost, PV(C_2), could then be expressed as an annual per-unit value by dividing its amortized (annual payment) value by the mean annual output as follows:

$$c_2 = \frac{PV(C_2)r\rho^T/(\rho^T - 1)}{\sum_{t=0}^{T} Q_t/(T+1)}. \tag{9.2}$$

The following discussions assume that per-unit values are available.

9.1.2 Satisfying the Three Conditions

Each of the three conditions (Table 9.2) presents challenges to the analyst. As will be seen, Conditions 1 and 3 are unlikely to be fully met, necessitating important judgment calls in deciding if, or how, to apply the method.

9.1.2.1 Condition 1

The first condition for applying the replacement cost method assures that the alternative provides the same benefit (i.e., the same service or a service of equivalent benefit) as the option being evaluated. The method relies critically on the existence of such a match. Quantity, quality, location, and timing are all important aspects of the service. If the alternative service is not identical to the proposed one, then some basis for comparing, say, different qualities, must be available.

In cases where only a less-than-perfect substitute is available, there are four possible situations (Steiner 1965). Assume, as before, that c_1 and c_2 are the unit costs of the public project and private alternative, respectively, and assume that the third condition, $b_2 \geq c_2$, can be accepted. First, if the quality of the output of the public project exceeds that of the private alternative and $c_2 > c_1$, we conclude that $b_1 > c_2$ (and also that $b_1 > c_1$). Second, if the quality of the output of the public project exceeds that of the private alternative but $c_1 > c_2$, we know that $b_1 > c_2$, but we cannot conclude that $b_1 \geq c_1$. Notice that in these first two situations we have determined a lower bound on b_1 but not measured it. In the third and fourth situations, the quality of the output of the private alternative exceeds that of the public project; the situations differ in the cost relations (either $c_2 > c_1$ or $c_1 > c_2$). These situations do not allow direct conclusions about the relation of b_1 to c_2.

Finding a perfect substitute is unlikely, and finding a close substitute is more difficult with some services than others. For example, electricity from a hydroelectric plant is interchangeable with electricity from a gas turbine plant, but options for shipping bulk goods can differ markedly; although both a barge and a freight train can move bulk goods, these options may differ in transit time and in the convenience of the origination and destination points. Verifying that Condition 1 is met is likely to be even more challenging for environmental services (e.g., water purification provided by a wetland) than for economic services that can more easily be commoditized (e.g., electricity, shipping of bulk goods) because of the complexities of environmental resources. To understand, for example, the water quality maintenance provided by a wetland well enough to know just what would be required of a perfect substitute is a tall order (Barbier 2011). In large part, when dealing with complex environmental resources, the specification of the service provided by Option 1 and the identification of alternative ways to provide that service (Option 2) are less an economic question than a matter of biological and physical science.

Satisfying Condition 1 becomes yet more difficult when not one but a whole set of services is at issue and when the alternative (Option 2) is not a built structure but rather another environment asset. For example, consider the valuation of the full set of services offered by a natural wetland. Perhaps the least-cost alternative way to replace that set of services is with a constructed wetland. Although it is possible for a constructed wetland to provide the same water quality protection as the original wetland, it is probably optimistic to think that a constructed wetland could match all the services provided by the original wetland, especially the passive use (i.e., nonuse) value of the original. As Ledoux and Turner (2002, p. 590) stated, "Cultural and historical aspects as well as a desire for "authenticity" may limit the extent to which nonuse values can be "transferred" in this manner to newer versions of the original. This is in addition to spatial and temporal complexities involved in the physical location of the new wetland or the time frame for restoration."

When valuing only a single service from an area that provides multiple services, other benefits and costs will go unaccounted for. That is, the two options are likely to differ in other ways that affect their overall benefits. For example, a hydroelectric dam and a thermoelectric plant may both provide electricity at desired levels and times, but they differ in their environmental impacts. Or, wetland preservation could provide similar water quality protection service to that of a water filtration plant, but valuing only that service would fail to capture other values of natural wetland preservation, such as the values of wildlife habitat or scenic quality. Thus, although the RCM may provide an estimate of the benefit of the service of interest, other valuation efforts could be needed in order to provide the information needed to make a fully informed decision about Option 1. Indeed, the viability of the option may hinge on doing so because it is feasible for measured b_1 to be lower than c_1, but for the full benefit of Option 1 to exceed c_1.

9.1.2.2 Condition 2

The second condition—the alternative provides a least-cost substitute—assures that the value of b_1 is not being overstated. It reflects the principle that we would not pay more for something than we have to. For example, if there are two candidates for Option 2 and one is more expensive than the other, we would insist on the lower-cost candidate. If there are several candidates for the least-cost alternative, costs of all of them must be estimated with sufficient care that they can be ranked. And if measurement of all costs is not feasible, it is necessary to provide evidence that unmeasured costs, if they were to be included, would not result in a different choice.

It perhaps goes without saying that the replacement cost method also relies on the replacement cost being determined in a relatively competitive market. Where the replacement cost is unduly affected by market interference or market failure, adjustments are necessary. For example, both electric rates and rail shipping rates may be subject to regulations intended to compensate for monopolistic advantages. Regulation is complicated and subject to political pressures such that rates may not reflect the social cost of providing electricity or moving freight.

Many public investments, and certainly the kind of investments amenable to use of the RCM, provide benefits for many years. In using the RCM to evaluate such investments, the analyst must understand how the cost of the alternative is likely to change over time. Two sources of change seem particularly important: technological change is likely to reduce future costs (Haveman 1972) and future environmental protection measures could increase the cost of the alternative.

Option 2 may impose external costs—costs not captured in the price consumers would pay for the service provided by Option 2. For example, a thermoelectric plant lowers air quality and releases carbon, resulting in costs that are not often reflected in the price consumers pay for electricity. Measuring and including those costs as part of c_2 might inappropriately increase the estimate of b_1 obtained using the RCM. The measure b_1 is an estimate of the service provided, electricity in this example. Unless we have good evidence that consumers would purchase the electricity at a price that includes the external costs, and unless the thermoelectric plant would (with the external costs included) still be the least-cost alternative option, we cannot claim that the service provided is worth the expanded cost. However, the external costs are relevant to the ultimate decision about implementing Option 1, if implementing Option 1 would allow the external costs to be avoided (as building the hydroelectric plant would avoid the external costs of the thermoelectric plant). Thus, the external costs of Option 2 are part of the costs mentioned above that go unaccounted for when using the RCM to estimate the value of a service, and that should be considered when deciding about providing Option 1.

9.1.2.3 Condition 3

The third condition—$b_2 \geq c_2$, which indicates that in the absence of Option 1, the alternative would be demanded if provided at cost c_2—assures that given Condition 1, Option 1 is worth as much as it would cost to replace it. If Condition 3 is not met, b_1 could be less than c_2.

Observed private provision of the alternative, as in the shipping example mentioned previously, offers the greatest assurance that the alternative would in fact be provided at a cost no more than consumers would be willing to pay. However, situations can arise where private provision is likely but not assured. For example, perhaps the private railroad that would ship goods in the absence of a public river dredging project is still under construction or is in operation but experiencing financial difficulties. Such uncertainties complicate application of the RCM and necessitate judgment calls by the analyst. Of course, such calls can be supported by sensitivity analysis.

Use of the RCM becomes even less clear-cut if the alternative would be provided by a public entity, as was recognized by Eckstein (1958, p. 70), who asked rhetorically if the benefit of a federal dam was limited by the cost of the federal dikes it would make unnecessary. He rejected use of the RCM in such situations because, although public provision of the alternative may suggest that $b_2 \geq c_2$, it does not assure it (see also Herfindahl and Kneese, 1974, and Young and Gray, 1972). The argument is essentially that if the agency controls the decision about whether or not to undertake either project, it cannot assert that the alternative (the dikes in this example) would definitely otherwise be undertaken and that such an undertaking would indicate $b_2 \geq c_2$. Furthermore, if the alternative would be provided by the same entity (and therefore out of the same budget) as the proposed project, the cost of the alternative would not be released for other purposes, thereby obviating part of the rationale for accepting the alternative cost as a measure of benefit. In this situation, assurance of $b_2 \geq c_2$ could only be provided by directly estimating the benefit of the alternative (b_2) and finding that it exceeded the cost

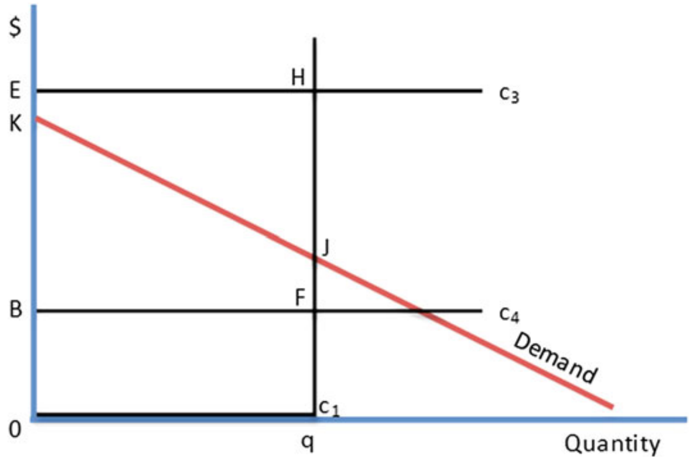


Fig. 9.2 Replacement cost possibilities in the absence of a viable alternative

(c_2), but if such estimation were feasible (at reasonable cost), we would not contemplate using the RCM.

Use of the RCM is also less clear-cut when the method is used to value a service, such as an existing ecosystem service, for which substitute provision cannot be observed. To explore this situation, consider Fig. 9.2 and assume that the ecosystem service is currently freely available up to quantity q (indicated by the marginal cost curve c_1), and that demand for the service in the absence of a substitute is given by the demand curve, which is unknown. If the least-cost way of providing a substitute is shown by marginal cost curve c_3, the replacement cost is area 0EHq. This area is larger than the actual total value of the service in the absence of a substitute, which is equal to area 0KJq. Given c_3, none of the service would be demanded, and c_3 would clearly overestimate the unit benefit. Alternatively, assume the substitute were available at marginal cost c_4. In this case, the replacement cost is area 0BFq, which is smaller than the value of the service in the absence of a substitute (0KJq). With a substitute available at cost c_4, the demand curve is bounded by c_4, and at least q of the service could be valued at c_4 per unit. Clearly, if demand cannot be estimated and observation of willingness to pay a certain positive cost (such as c_4) is impossible, either of these possibilities could exist (as could many others), and use of the RCM to estimate the benefit of an ecosystem service requires some convincing evidence that Condition 3 is met.

However, as use of the RCM has been extended in recent decades to valuation of nonexclusive environmental services for which the alternative is unlikely to be provided by the private sector, reconsideration of the initial restrictive Condition 3 has become necessary.[7] To examine the situation more carefully, let Option 1 be the

[7]de Groot et al. (2012) reported that the replacement cost method (and the avoided cost method) have been the most commonly used methods for valuation of regulating ecosystem services.

protected environment and Option 2 be an alternative source of services comparable to those of the protected environment. If we can accept that Condition 1 is met but we lack full confidence that Condition 3 (in the absence of Option 1, $b_2 \geq c_2$) is satisfied, we are left with b_1 being limited by c_2 (because we would not be willing to pay more for something than the cost of a perfect substitute) but not necessarily equal to c_2 (i.e., $b_1 \leq c_2$). Thus, in comparison with the situation where Option 2 would be provided by the private sector, applications where the alternative would be publicly provided are less likely to provide an estimate of b_1 ($b_1 = c_2$) and more likely to merely provide an upper bound on b_1 ($b_1 \leq c_2$).

The role of a legislative mandate. The RCM is sometimes used to value environmental services that could be lost if not protected but which by law would be replaced if lost. To explore this case completely, we ask: if there are two substantially different options for achieving the same goal, and the second option (Option 2) is legislatively mandated and will go forward unless the first option (Option 1) is implemented, can the cost of the second option (c_2) indicate the value of the first option (b_1)? We know that, given Condition 1 is satisfied, we would not pay more than c_2 per unit for Option 1, so we affirm that $b_1 \leq c_2$. But to be confident that $b_1 = c_2$, we must also know that in the absence of Option 1 $b_2 \geq c_2$. The passage of the legislation requiring that the goal be met is supportive of $b_2 \geq c_2$, but because it provides no assurance that $b_2 \geq c_2$ holds in this specific case, Condition 3 is not necessarily satisfied, and we are left knowing only that $b_1 \leq c_2$.

Interestingly, in the presence of a mandated goal, a measure of benefit is not needed by the agency proposing the project. If the goal must be met, only the costs of the options matter to the agency, and the agency's decision is simply one of computing and comparing aggregate costs of the two options and choosing the least expensive way of achieving the goal assuming equality of outcomes. This conclusion is valid as long as it is clear that the alternative cost would be borne if the project did not go forward. Of course, the question of whether or not the mandated target enhances public welfare is another matter.

New York City's acclaimed decision to protect the watershed where its water supply originates demonstrates this point. For its Catskill-Delaware watershed, the city chose watershed protection rather than adding filtration to the city's water treatment system (Committee to Review the New York City Watershed Management Strategy 2000). Regulations for implementing the Clean Water Act in the United States require that drinking water reach a certain level of purity, and it was determined that some additional action would be needed for the city to remain within the regulated levels for certain pollutants. Let Option 1 be watershed protection and Option 2 be construction of a filtration plant. Option 1 was found to be much less expensive than Option 2. Although there was some uncertainty about the future success of the watershed protection option, it was essentially concluded that both options would provide water meeting Clean Water Act requirements, thus satisfying Condition 1. A filtration plant was the standard option in such cases (most cities in the United States have one), and its rough cost could be determined, so with careful study Condition 2 could be satisfied. We have no direct estimate of

b_1, or of b_2, but we know that, because of the Clean Water Act, the alternative cost will be borne in the absence of Option 1. If we cannot accept that the existence of the law verifies that $b_2 \geq c_2$ and lack other evidence supporting that case, we can conclude only that b_1 is the lesser of the cost of a perfect substitute and a directly measured estimate of b_1, whatever that is, thus, $b_1 \leq c_2$. (It could be argued, though, that in obvious matters of human health such as drinking water safety, the public is likely to consider the benefit worth the cost.) However, an estimate of b_1 is not needed for the city to make a decision. When one or the other option will surely be implemented, the city's decision is resolved simply by comparing costs.

Levels of confidence. Differences among the ways alternatives can be provided are perhaps best considered as a matter of the confidence with which Condition 3 is satisfied. The greatest level of confidence, as already mentioned, is offered where the alternative definitely—or at least very probably—would be provided by the private sector. The second level of confidence occurs when the alternative would be provided by a different public entity from the one that would provide Option 1—one that is free to act in the interests of its constituents. For example, construction by a federal agency of a flood control dam would preclude the cost of emergency flood response and rebuilding that a local government entity would otherwise incur. Steiner (1965) called options satisfying these levels of confidence "viable alternatives." Such provision would suggest that $b_2 \geq c_2$.

The third level of confidence occurs where a viable alternative is not available but the alternative would nevertheless certainly be provided. The most likely source of this certainty is a law requiring that some target be met. For example, consider the case (mentioned previously) where a federal law requires that water providers meet certain drinking water quality standards. Passage of the law by the legislators, as the voice of the public, indicates a willingness to pay the cost of reaching the standard, whatever it is. Now suppose a public land management agency was considering a vegetation management project to lower erosion from its lands, which would have the effect of precluding the need for the downstream city to improve its drinking water treatment plant to ensure that its water meets the standard. In considering use of the RCM to value the water quality protection its project would provide, the agency could not in full confidence assume that the city's action to improve the treatment plant was independent of the federal law (and thus that the cost of the improvement would indicate the benefit of the vegetation management project), but passage of the law would provide some support for assuming that the cost of the improvement to the plant was considered worth the benefit.

The lowest level of confidence occurs where there is no direct indication that the alternative would otherwise be provided, but there is some supporting evidence for accepting $b_2 \geq c_2$. That evidence could include observation of other entities undertaking a project like Option 2 under no coercion to do so or some benefit transfer evidence that supported accepting $b_2 \geq c_2$.[8]

[8]A completely unsupported use of RCM occurs where the alternative would be provided by the same entity as that proposing the new public project, and the entity simply offers as the alternative

Having defined four levels of confidence that Condition 3 is met, we now ask how much confidence is enough, assuming that the first two conditions in Table 9.2 are met. Eckstein and Steiner required levels 1 or 2. More recent writers (e.g., Young 2005) have allowed for lower levels but without being specific about the nature or sufficiency of the required corroborating evidence. In the end, the analyst must provide the answer on a case-by-case basis, knowing that any level below 2 is subject to question. Here again, sensitivity analysis should be helpful.

It is interesting to note that where the RCM can be applied most confidently is where the service to be provided by the public project would otherwise surely be provided by the private sector. The most pertinent question may not be, "What is the value of the public project?" but rather, "Can the public project provide the service at a lower cost than the private sector?" A comparison of costs indicates whether and by how much the public project would increase public welfare. Thus, a measure of the benefit of the public project may be unnecessary in the situation where the method is most confidently applied—all that may be required is a cost-efficiency analysis.

An important distinction between the RCM and the methods described in other chapters of this text is that the RCM relies principally on data about the behavior of firms (producers) or public entities, whereas the other methods use data about individual consumer behavior (or statements about that behavior). Individual consumer behavior is presumed to provide more direct and viable information about willingness to pay and is generally preferred by economists as the source of value estimates (e.g., see National Research Council, 2005). However, the importance of this concern depends on the application to which the method is put. When the RCM is used to value a public effort to provide a service that would otherwise surely be provided by private firms, the firms typically would be acting in response to individual consumer desires. It is when the alternative would not otherwise be provided by the private sector—and thus when confidence is lower that Condition 3 has been satisfied—that concerns about the RCM are more relevant.

9.1.3 Steps in Applying the Replacement Cost Method

Having described the essential conditions for applying the RCM, the steps for applying the method can be summarized (Table 9.3). Step 1 requires an initial assessment of whether or not the requirements for applying the method are likely to

(Footnote 8 continued)

a slightly more expensive project to provide the service, thereby assuring a cost savings if the proposed project were undertaken. This possibility led to the suggestion that the alternative must be provided by a substantially different means from that of the proposed project (Herfindahl and Kneese 1974; Young and Gray 1972). However, the "substantially different" requirement does not assure that Condition 3 is met.

Table 9.3 Steps of RCM (resource protection language in brackets)

Step 1	**Assess the options and set basic parameters of the analysis**
	• Identify Option 1 and one or more possible alternatives (Option 2)
	• Consider the likelihood that the three conditions in Table 9.2 can be satisfied (if likelihood is low, consider other valuation options or whether valuation is actually needed or appropriate)
	• Specify the planning time horizon (T) and the discount rate (r)
Step 2	**Specify Option 1**
	• Define the service to be provided [that could be degraded or lost] and the proposal (Option 1) to provide [protect] the service
	• Estimate amounts and qualities of the output that would be made available [protected] over the time horizon
	• Determine the users of Option 1 and specify the part of the total output that would actually be consumed by [protected for] those users
	• Estimate the cost of Option 1 (c_1), including (a) list actions (and related inputs) and when they would be performed over the entire planning horizon, (b) estimate the annual cost of all actions, making adjustments for price effects and market imperfections, and (c) compute the present value of the costs
Step 3	**Specify Option 2**
	• Specify alternatives for Option 2 that satisfy Condition 1 to a reasonable degree
	• Estimate the costs of each alternative for Option 2 over the identified time horizon (see Step 2 for details)
	• Select the alternative for Option 2 with the lowest cost (Condition 2)
Step 4	**Settle on the replacement cost**
	• Compare cost estimates to ensure that $c_1 \leq c_2$ (Condition 2)
	• Affirm that Condition 3 is satisfied to a reasonable degree; consider the level of confidence that Condition 3 is satisfied and whether it can be concluded that $b_1 = c_2$ or just that $b_1 \leq c_2$
	• Isolate any differences in amount and quality between the outputs of the two options and adjust the cost estimate for differences in the services provided (e.g., in timing, reliability, convenience) by the two options
Step 5	**Document ancillary benefits and costs of the options (benefits and costs that are separate from the service at issue and, therefore, not reflected in the replacement cost)**

be satisfied to a sufficient degree. If they are, then basic parameters of the analysis are specified—most importantly the planning horizon and discount rate.

In Step 2, the proposed measure (Option 1) is carefully described and its cost is estimated. First, the service to be provided (or protected) is defined, and the proposed project is outlined in detail, including specification of the output amount and quality of the service over the planning horizon. Estimating the amounts and qualities of the output can involve understanding complex physical and biological processes. Also, a distinction may be necessary between the potential output level and the amount of output that would actually be consumed over the planning horizon, for it is the latter that is relevant to the analysis, which requires projections of future demand.

Once the project is carefully outlined, the cost is estimated. Cost estimation is generally less involved than benefit estimation, but it is potentially quite challenging. In its simplest form, the tasks to be performed and the materials, land, etc., to be purchased are listed and then the purchase costs of those inputs are obtained or estimated. Standard unit cost estimates could be available to facilitate the task. Three basic assumptions in using existing market prices to estimate costs are as follows: the project's use of the inputs will have minimal impact on other uses of the inputs and, therefore, will not affect input prices; the real prices of the inputs will not change over time due to influences unrelated to the proposed project; and the prices reflect the true opportunity cost of the inputs.

The larger the project, the greater is the likelihood that its demand for inputs will affect market prices. However, it is probably safe to expect that, except for large infrastructure projects, the first assumption will not be seriously violated. The second assumption is that technological or other changes will not lead to a significant change in prices of the inputs within the time it takes to implement the project. Projects that take only a few years to implement are likely to avoid this complication. The potentially most troublesome assumption is that market imperfections are not unduly affecting the market prices of inputs. Such imperfections could arise due to price subsidies or other price controls that interfere with what would otherwise be a relatively competitive market, market concentration (monopoly, oligopoly) controlling prices, and externalities in producing the inputs, which may impose costs (or award benefits) on third parties.[9] Adjustments for such market imperfections, in order to estimate prices that reflect the true social opportunity costs of the inputs, as well as adjustments for nonmarginal effects, require sophisticated analysis. Such analysis is essentially the same as what would be necessary to correctly measure the benefits of a public program that would allow an increase in production of the inputs.

Large projects may be implemented over several years, in which case discounting will be needed to produce a cost estimate that can be compared with the cost of Option 2. The present value of the total cost would then be estimated as in Eq. (9.1). If inflation-adjusted costs have been estimated, the discount rate (r) should reflect that adjustment.

In Step 3, the alternative (Option 2) for providing the service at issue is specified and its cost is estimated. Several options could be considered so as to end up with the lower-cost option that provides a service matching (as closely as is reasonably feasible) that of Option 1. As with Step 2, concerns about the effects of market imperfections could be relevant.

[9]An additional concern is that labor cost does not reflect the true opportunity cost because of lack of full employment of labor. One way to account for this is to estimate the labor cost as its nominal cost times $(1 - p)$, where p is the probability that use of the labor will have no net effect on use of labor elsewhere.

Then in Step 4 the estimated costs of the two options are compared to make sure that Condition 2 ($c_1 \leq c_2$) is satisfied. If Option 2 is not a privately provided service, Condition 3 ($b_2 \geq c_2$) would not be met with a high level of confidence, which should prompt consideration of the possibility that the analysis cannot conclude that $b_1 = c_2$. Further, appropriate adjustments are made for any major differences between the services provided by the two options.

Finally, in Step 5, benefits and costs of the two options that are separate from the service being valued must be documented in order to provide a full picture of the project being considered.

9.1.4 Applications of the Replacement Cost Method

Four examples of RCM applications are described below. But before proceeding with the examples, it is important to note that not all studies that estimate replacement costs are RCM applications. First, replacement costs are often estimated in other contexts, and such estimations should not be confused with application of the RCM. For example, a recent study estimated the cost of replacing via commercial fertilizer the major nutrients (nitrogen, phosphorous, and potassium) that would be lost from the soil if crop residues were removed from farm sites (to be used to produce cellulosic ethanol; Chatterjee 2013). Although commercial fertilizer is a least-cost and nearly perfect substitute for the lost nutrients and could almost certainly be relied on if necessary—thus satisfying the three conditions in Table 9.2—the computation of replacement cost in this case is not an application of the RCM because the residue is a private good, and the study objective was simply to estimate a replacement cost, not to value a public resource or expenditure.[10]

Second, the RCM, as well as other methods, are sometimes used to estimate the value of a naturally occurring service that is unlikely to ever be completely lost. Although such a value is probably of no practical significance, it may be of interest as a way to demonstrate the importance of natural ecosystems. For example, Hein (2011) valued the groundwater infiltration that occurs in a protected forest as the additional cost of water treatment if the groundwater were no longer available, and in its place, water would be diverted from a source of lower quality. This application probably satisfies the three conditions for application of the RCM (Table 9.2), and thus the replacement cost gives the value of the water quality increment provided by groundwater that is currently pumped within the protected forest. However, because there is little chance that water infiltration to and filtration within the aquifer would be completely lost if the forest was not protected, the value is of a change that could not happen and, thus, it does not answer a realistic management question.

[10]This is not to suggest that the estimate of replacement cost is not socially relevant. Because the residue would be removed from farms to produce cellulosic ethanol to meet a target established by a U.S. federal law, the Energy Independence and Security Act of 2007, the cost of the commercial fertilizer is one part of the full cost of responding to the Act.

Of the four RCM examples presented next, the first two are of possible public investments to provide a new source of a particular service, and the other two are of possible public protection of an existing service. These examples highlight some of the challenges and constraints that are faced in implementing the RCM.

9.1.4.1 Electricity from a Hydroelectric Dam

We begin with a relatively straightforward application of the RCM—valuation of electricity that could be produced at a proposed public hydroelectric dam. Before beginning to work through the steps for application of the RCM, however, it is reasonable to ask if another valuation method would be more appropriate. Perhaps the most obvious alternative method would be to simply value the electricity at the price at which electricity in the area currently sells and is likely to sell in the future. This approach would be appropriate as long as the electricity were sold in a competitive market—that is, one with numerous producers and without market interference from government regulations. Because electricity producers in an area tend to be few and electricity prices are typically regulated, use of existing prices is problematic and the RCM remains a viable candidate.

Beginning with Step 1 (Table 9.3), use of the RCM is likely to satisfy the three conditions for an RCM application (Table 9.2). Electricity from different sources, given certain timing considerations, is interchangeable, which allows electricity from different sources to be combined in a large grid for distribution to customers. Thus, electricity produced at a hydroelectric plant substitutes for electricity produced in other ways, satisfying Condition 1. The next least expensive alternative to a hydroelectric plant is most often a privately run thermoelectric plant (usually a coal- or natural-gas-fired plant), satisfying Condition 2. If the electricity from the hydroelectric plant would in fact otherwise be produced at thermoelectric plants, satisfying Condition 3, the gross unit benefit of the hydropower is equal to the cost of producing the electricity at selected thermoelectric plants (c_2).[11]

When the requirements of Step 1 are satisfied, Step 2 begins with specification of the electric output of the proposed hydroelectric plant and determination of the amount of that output that would be consumed over the planning horizon (which could be the life of the plant or some shorter time). If the hydroelectric plant would have excess capacity, only the portion expected to be used would be valued at the alternative cost. Estimating the amount to be used would require projecting electricity demand over the planning horizon in light of other expected sources of supply. Specification of the plant output would involve considering the likely future streamflow that could be diverted through the plant and the timing of those diversions, which would depend in part on inflow to the plant and any multiple use

[11]The water input in hydropower production is typically available without cost to the hydropower facility. The unit cost savings from construction of the hydropower plant, $c_2 - c_2$, can be attributed to the availability of the water input, and can be considered the value of the water in hydropower production.

considerations (e.g., flood control, recreation) affecting management of the reservoir upstream of the plant, as well as any in-stream flow considerations (e.g., protection of fish habitat) affecting management of reservoir releases. Inflow to the plant would be subject to hydrologic fluctuation, including drought. These questions would require consideration of any laws and standing agreements affecting water management in the area. Most likely, a sophisticated river basin water management model would be used to simulate options for management of the proposed hydroelectric facility. To complete Step 2, the costs of constructing, managing, and maintaining the hydroelectric plant and related facilities would be estimated, with adjustments made for expected price effects or market imperfections.

In Step 3, Option 2 is defined in light of the three conditions (Table 9.2). Regarding Condition 1 (finding an appropriate substitute for the electricity to be produced at the hydroelectric plant), a key issue is the timing of the electric output. In the Western United States, for example, it is common for output at hydroelectric plants to be ramped up or down by changing the amount of water that is run through the turbines, allowing production to quickly respond to changing real-time demand. Production at peak demand times is more valuable than production at other (base load) times. Thus, the thermoelectric plant costs that are used as the estimate of c_2 should be taken from plants, such as gas turbine plants, that provide electricity with comparable flexibility in timing to what would be produced at the proposed hydroelectric plant, rather than from plants providing only base load power, like coal-fired plants.

Next the replacement cost is estimated. The following are two issues that would arise in measuring the cost and assuring that the cost is the least-cost way of providing the alternative (per Condition 2). First, existing capacity at thermoelectric plants would need to be estimated. If there were sufficient capacity at existing thermoelectric plants to accommodate demand over the planning horizon if the hydroelectric plant were not built, only variable costs at the thermoelectric plants need be taken into account. However, if existing thermoelectric plants would not be able to fully meet expected demand, new capital costs would also need to be included in the estimate of c_2. Second, variable costs at thermoelectric plants are dominated by the cost of fuel. As recent swings in natural gas prices in the United States demonstrate, these costs are difficult to predict, but informed estimates on this issue are critical to arriving at a realistic estimate of the least cost. In addition, the cost estimate would be adjusted to account for market imperfections affecting the alternative cost and for differences between services provided by the two options (including those introduced by constraints on management of the hydroelectricity facility, as mentioned previously).

In Step 4, the costs (c_1 and c_2) are compared to verify that the public project would be an improvement, and then remaining issues are addressed. In the case of hydroelectric power, it is likely to conclude that $b_1 = c_2$.

Finally, in Step 5 ancillary benefits and costs would be documented. These could include the recreation that would occur in the new reservoir created by the dam and the reduction in downstream recreation that would result if the dam affected fish habitat or other in-stream uses, the loss of use of the upstream riparian area that would be inundated, the flood control that the new dam would provide, and the reduction in air pollution (including greenhouse gas emissions) that substitution of hydroelectricity for thermoelectricity would allow.

9.1.4.2 Coastline Protection via Planting Mangroves

Many mangrove forests along the coast of Vietnam and elsewhere have been lost to construction projects, aquaculture, and other changes. As mangroves have been removed, the services that they provide have been more widely recognized, leading to efforts to replant mangroves. RCM is among the methods commonly used to value services provided by mangroves (Brander et al. 2012). A good example was provided by Adger et al. (1997), who described a proposal by an international development agency to plant mangroves in an area of northern Vietnam where a sea dike had been used for centuries to help protect nearby intensively used farmland from coastal storm surges and floods. A mangrove forest in front of the dike had once trapped sediment and dissipated wave energy, and loss of the mangroves resulted in increased costs to maintain the dike. The service at issue is farmland protection from flooding. The study valued the proposed mangrove forest (Option 1) as the avoided cost of maintaining the dike (Option 2) and used a model expressing dike maintenance as a function of wave length and the width and age of the future mangrove stand. Condition 1 is met as long as we count only the amount of dike maintenance that is avoided by having the mangrove forest. Condition 2 is satisfied if the dike was being maintained at least cost, which is likely, and if those costs can be accurately measured. Further, it can be accepted that Condition 3 is met because the local people were already maintaining the dike. Focusing only on the flood control benefit of the mangroves and knowing therefore that the willingness to pay for the mangroves is limited by the cost of the alternative, it could be concluded that $b_1 = c_2$.[12]

This case is a rather uncomplicated application of the RCM. What it most importantly demonstrates is the danger of focusing on only one benefit of a proposed project. Lowering dike maintenance costs is only one of several benefits of a mangrove forest. In fact, the mangrove option was found to be more expensive than

[12] In the coastline protection example, use of the RCM is related to use of the defensive behavior method (see Chapter 8). Recall that the defensive behavior method equates the cost of an individual's defensive measure with the benefit to the individual of a policy that would avoid the need for defensive expenditure, on the assumption that the individual would be willing to pay at least the cost of the defensive measure to avoid the damage. The mangrove example equates the cost of additional dike repair (which avoids damage to crops and other investments) with the benefit of a policy (the mangrove forest) that would avoid the need for the defensive measure, on the assumption that society would be willing to pay at least the cost of the defensive measure to avoid the need for the measure. Defensive measures have been observed in both cases. The situations differ in that with the defensive expenditure method we observe individual defensive measures being taken, whereas in the mangrove case we observe a community-wide defensive measure—maintenance of the dikes.

dike maintenance ($c_1 > c_2$), such that $b_1 < c_1$. However, when the values of other benefits—fish habitat, wood, and nectar for honey production—were added to b_1, the study found that the measure of total benefits exceeded the cost of the mangrove project.

9.1.4.3 Flood Avoidance via Floodplain Protection

Natural floodplains soak up water, thereby limiting damage that would otherwise occur within the floodplain in times of unusually high precipitation. Floodplains are continually under development pressure and, thus, an estimate of the value of protecting a natural floodplain could serve a useful purpose. Use of the RCM might be appropriate here, where Option 1 is floodplain protection and Option 2 is flood avoidance using an alternative approach that allows for land development, such as infrastructure construction (levees, dams, etc.). But before beginning an RCM application, in this case one would need assurance that the avoided damage cost achievable with the options exceeded the implementation cost, for clearly if the cost of avoiding the damage would exceed the cost of the damage, the most efficient course would be to endure the damage. To address this issue, the amount of flooding that would occur would need to be determined with specified probability, with and without the protection options over the planning horizon selected for the analysis, and then the damage that would occur in both the "with" and "without" cases would be estimated. This estimation would involve complex meteorological and hydrological analysis plus assessment of the physical effects of flooding on outdoor and built environments because those environments would be expected to evolve over time. (Notice that this effort would go a long way toward providing a direct estimate of the benefit of flood protection—that being an estimate of the avoided damage cost—perhaps obviating the need for an RCM application.)

Proceeding with the RCM analysis, a review of the three conditions for use of the RCM (Table 9.2) would conclude that the use of this method would be viable as long as the infrastructure could provide the same flood protection as the natural floodplain, the infrastructure was the least-cost alternative way to provide that service, and the public would build the infrastructure if the flood protection services of the natural floodplain were lost. The comparison of cost with damage avoided, mentioned above, would provide an answer to the last of these concerns.

Step 2 requires specifying the details of floodplain protection. The protection plan might include purchase of land in the floodplain, zoning changes, and construction of detention ponds, among other measures. Then the cost of the plan would be estimated, with particular attention to estimating the cost of land purchase (which would reflect the value of the land in alternative uses). Moving to Step 3, Option 2—the least-cost alternative to floodplain protection—would be defined. Arriving at the best combination of available measures would require substantial study. A complicating factor in designing an Option 2 that satisfies Condition 1 is that floodplain protection and infrastructure construction may not be equally effective at each different size flood. Then the cost of Option 2 over the time horizon would be estimated in light of market imperfections that could affect the cost estimates.

In Step 4, differences in the flood protection (and thus in damage) that would occur with the two options would be considered. These differences might include the maintenance costs over time, the relative capacities of the two approaches to control flooding in light of projected increases in flooding under climate change, and the scenic impacts of the alternatives. Finally, the cost estimates are adjusted for any differences in the services provided by the two options.

As nearly always in economic valuation, the particular application may ignore ancillary benefits or costs, which are examined in Step 5. In the flood protection example, both options, floodplain protection and infrastructure construction, have implications for what happens in the floodplain. Floodplain protection allows some activities (e.g., a public park, waterfowl habitat), whereas infrastructure allows others (e.g., farming, urban development). Although these other benefits are outside the purview of the valuation effort—and indeed tend to be forgotten as attention is focused on the service being valued—they should also be considered when choosing between the options.

9.1.4.4 Food Traditionally Harvested by Indigenous People

What is the value of subsistence harvest of wildlife? Jackson et al. (2014) used the RCM to obtain such a value. They estimated the amounts of various aquatic species harvested for food by indigenous peoples in three large river systems of Northern Australia and then matched each species with commercially available food products and used the retail costs of the products as estimates of the value by weight of the harvested species. For example, long-necked turtle was matched with T-bone steak, and pig-nosed turtle with filet steak. The replacement cost per household per fortnight for each species was then estimated based on the amount of the species typically consumed and the supermarket price of the replacement product.

In this application, Option 1 was essentially maintenance of the habitat where the species live, and Option 2 was reliance on commercially available products. Conditions 1 and 2 (Table 9.2) are reasonably well met in that the authors attempted to match each species with a food product of similar quality and were careful to use the lowest available prices. However, condition 3 is most probably not met, in that the subsistence users tended to be short of cash and relied on gathering in part to avoid the high cost of comparable commercial products (this introduces equity concerns that are not addressed by the valuation exercise). Although failure to satisfy the third condition with much confidence complicates any conclusions that could be drawn from the study, the application nevertheless serves at least two useful purposes: it demonstrates that the subsistence harvest has substantial value and shows which species are of highest value and, therefore, which habitats should be of higher priority for protection.

It should be noted that, like other studies, this one attempts to measure only part of the full value of the respective activities. The subsistence harvest study focuses on the food value of the harvested species. Other values, such as the value of the harvest activity itself—which could well exceed the food value—would be estimated using other methods.

9.2 Equivalency Analysis

Equivalency analysis (EA) is not a valuation method, but it is a substitution method, and in one form it employs economic valuation. EA is used to identify the actions needed to provide replacement resources or services equivalent to resources or services that have been or will be lost. The end product of EA is an estimate of the cost of performing those actions.

Equivalency analysis has three variants: resource equivalency analysis (REA), habitat equivalency analysis (HEA), and value equivalency analysis (VEA). To determine the actions needed to replace lost resources or services, REA focuses on the resources themselves (e.g., number of lost salmon) and HEA focuses on the services that flow from land or water areas (e.g., water quality protection), without attention to the economic values of the resources or services. In doing so, EA is said to use resource-to-resource or service-to-service scaling. In contrast, VEA relies on measures of economic value and is said to use value-to-value scaling (Chapman et al. 1998; English et al. 2009).[13]

Basically, the EA approach is to characterize a loss, decide on a way to provide resources or services equivalent to those lost, and then "scale" the replacement activity—that is, decide on its size or extent—so that the total gain is equivalent to the total loss. The "service-to-service scaling" terminology, for example, indicates that if a given quantity of a certain service is lost, a way will be found to provide a quantity of the same or a similar service so that the loss has been wholly offset.

EA arose in response to laws requiring compensation for lost resource services and has been used in conjunction with both future and past losses. A major ex ante application follows from the Clean Water Act in the United States, which adheres to a no-net-loss policy requiring that if construction projects will unavoidably cause a loss in wetland services, that loss must be mitigated by improving or creating other wetland areas.[14] Similar policies exist elsewhere.[15] Ex post applications of EA deal with unplanned resource losses, such as from offshore oil spills and leaching of

---

[13]HEA and REA came about partly in reaction to the cost of economic valuation—especially in situations where the damage is modest, the cost of analysis can be excessive in relation to the importance of the damage—and to court challenges of valuation efforts (Roach and Wade 2006).

[14]The earliest exposition of the REA approach to compensating for resource losses could be King and Adler's (1991) description of a process for determining compensation for anticipated wetland loss.

[15]For example, Canada's Fisheries Act contains a no-net-loss provision regarding the productive capacity of fish habitats that could allow for use of REA (Clarke and Bradford 2014; Quigley and Harper 2006).

toxic chemicals into groundwater and are a common component of natural resource damage assessments.

Natural resource damage assessment is the process of developing the public's claim against the parties responsible for natural resource damages. The discussion here is based on the United States experience, but similar laws and procedures apply elsewhere.[16] Natural resource damage assessment involves assessing the injuries to natural resources or services and then developing a plan to speed the recovery of injured resources or areas (called primary restoration) and compensate for losses incurred before recovery is complete (compensatory restoration). EA is an approach for the latter—for determining the actions to be taken to provide compensation for losses and estimating the cost of those actions.

The goal of natural resource damage assessment is to make the public "whole" for injury or loss of natural resources and services. Under the natural resource damage assessment framework, the public is made whole by receiving resources and services of comparable value to what was lost. One might assume that the most straightforward approach to compensation would be to pay the parties suffering losses,[17] but damage assessment regulations for implementing the Oil Pollution Act of 1990; the Comprehensive Environmental Resource, Compensation, and Liability Act of 1980 (CERCLA); and other federal statutes require public trustees to use compensatory funds only to "restore, rehabilitate, replace, or acquire the equivalent of the injured resources" (NOAA 2006, p. 2).[18] Given this, even if trustees computed the monetary value of the losses incurred and collected that amount from the responsible party, they would still need to spend the money on restoration and related projects, with no assurance that the resources or services received from the projects would equal the resources or services lost because of the damage. Early natural resource damage assessments performed pursuant to CERCLA and the Clean Water Act emphasized use of an economic valuation approach—essentially VEA—to determining equivalency,[19] but the Oil Pollution Act of 1990 expressed a preference for an approach—essentially HEA or REA—that shifts the goal from measuring the economic value of interim losses to estimating the cost of creating compensatory resources or services. This approach assures that the collected funds are sufficient to pay for the designated replacement project.

[16]The European Union's Environmental Liability Directive was established in 2004 and incorporated into EU member states' domestic law by 2010. Like comparable U.S. legislation, the ELD gives priority to physical/biological equivalency over value equivalency in the determination of compensatory liability as long as the replacement resources are similar in type, quality, and quantity to the lost resources.

[17]As Randall (1997) explained, the most efficient approach to compensation is to choose the less expensive of two options: reimburse the parties suffering the losses or provide equivalent services.

[18]The National Oceanic and Atmospheric Administration of the U.S. Department of Commerce serves as trustee for coastal and marine resources and determines the damage claims to be filed against parties responsible for certain injuries to natural resources.

[19]The assessment performed following the 1989 Exxon Valdez oil spill is perhaps the best known compensatory analysis using an economic valuation approach (Carson et al. 2003; Goldberg 1994).

CERCLA focuses on contaminants, and the Oil Pollution Act focuses on oil spills, but other statutes allowing for compensation for unplanned natural resource losses apply as well to some physical injuries, such as to coral reefs from vessel groundings (Viehman et al. 2009).[20] In addition, EA recently has been used to determine required compensatory payments for damage to natural resources following large human-caused forest fires (Hanson et al. 2013; Kimball 2009). Settlements included not only recovery of fire suppression costs and the value of lost timber harvest, but also compensation for lost environmental services.

9.2.1 Characteristics of Equivalency Analysis in an Ex Post Context

Figure 9.3a depicts, in terms of resource losses, the principal elements of resource damage and recovery. Following a damaging event, a baseline is determined, which is the expected level of resources if the damage had not occurred. As depicted in the figure, the baseline can vary over time. The damage lowers resource levels, which may recover naturally. Restoration efforts speed up the recovery but still leave interim losses. Those losses can then be offset via compensatory actions. The compensatory efforts are likely to occur at a different site with its own baseline resource level, where they raise the resource level at the site compared with what would otherwise occur. Figure 9.3b depicts the case where compensatory actions at the alternative site raise resource levels above a more or less constant baseline, but it is also possible for the compensatory actions to maintain an initial baseline resource level that would otherwise fall. The gains are counted until the last year of the planning horizon. Compensation is complete if the gains (represented by the shaded area in Fig. 9.3b equal the losses (represented by the shaded area in Fig. 9.3a. Obviously, the sooner that primary restoration returns resource levels to the baseline level, the smaller is the required compensatory restoration effort.

Figure 9.3 simplifies the situation in at least two ways. First, the figure ignores the discounting of gains and losses that is a key feature of EA. Discounting is used to place the gains and losses on a common temporal footing, as compensatory restoration likely happens after damages have occurred.[21] Second, Fig. 9.3a depicts full recovery within the planning horizon. If full recovery is not possible, the damage assessment would determine the necessary compensation for whatever losses are expected to persist throughout the planning horizon.

[20]Other U.S. federal statutes affecting natural resource damage assessments include the Clean Water Act, the Park System Resource Protection Act, and the National Marine Sanctuaries Act.

[21]Regulations for natural resource damage assessments require discounting: "When scaling a restoration action, … trustees must take account of the value of time in the scaling calculations by discounting to the present the interim lost services or the value of interim lost services due to the injury, as well as the gain in services or service value from the restoration action" (61 FR No. 4, p. 453).

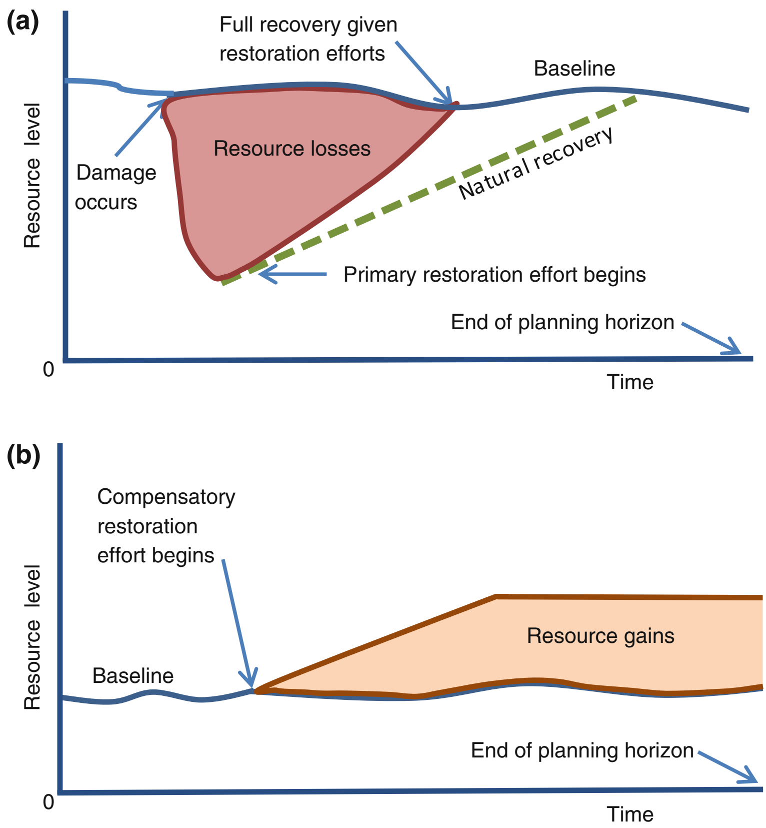


Fig. 9.3 Restoration for losses following damage. Primary restoration limits losses (**a**), and compensatory restoration provides gains at a separate site (**b**)

The EA approach to determining the necessary compensatory actions—which of course must be determined before the cost can be estimated—can be explained in terms of our earlier (RCM) framework using a variation of a simple model (Dunford et al. 2004) that captures the key components of the procedure for measuring losses or gains. That framework considers two options for yielding equivalent results. The model includes a value component, which facilitates highlighting the key difference between VEA and HEA or REA. Let Option 1 be the damaged area or resource and Option 2 be the replacement site where compensation for the losses will occur; thus "option," as used in the RCM framework, is replaced with "area" or "resource" in the EA framework.

The model is first explained in the context of REA, that is, in terms of resource losses (e.g., numbers of animals or gallons of water). The present value of the losses, V_1, is

$$V_1 = \sum_{t=0}^{T} v_{1,t}\left(\gamma_{1b,t} - \gamma_{1a,t}\right)\rho^{-t}, \tag{9.3}$$

where the damage occurs in year 0, $\gamma_{1b,t}$ is the baseline number of units expected in year t, $\gamma_{1a,t}$ is the actual number of units in year t, $v_{1,t}$ is the value of the average unit in year t, T is the last year of the planning horizon, and ρ^{-t} is the discount factor in year t ($\rho = 1 + r$, where r is the discount rate). Thus, the term in parentheses is the number of units lost in year t. If $\gamma_{1a,t}$ reaches $\gamma_{1b,t}$ within the planning horizon (i.e., if full recovery occurs before year T), the difference within parentheses goes to zero for all remaining years.

Two simplifying assumptions are often applied. First, the baseline service level ($\gamma_{1b,t}$) is assumed to remain constant over time at the γ_{1b} level.[22] Second, the average real (inflation-adjusted) unit value ($v_{1,t}$) is assumed to remain constant over the planning horizon. Indeed, $v_{1,t}$ will remain constant unless it is judged that extant conditions (e.g., increasing or decreasing scarcity of the resource) dictate otherwise. With these two simplifying conditions, V_1 becomes

$$V_1 = v_1 \sum_{t=0}^{T} \left(\gamma_{1b} - \gamma_{1a,t}\right)\rho^{-t}. \tag{9.4}$$

Knowing V_1, the present value of the lost resources (Eq. (9.3) or (9.4)), we now need to compute a similar level of resource value for Option 2, the replacement. As with the damage, the present value of the enhancement, V_2, is a net of two resource flows, as follows:

$$V_2 = \sum_{t=H}^{T} v_{2,t}\left(\gamma_{2a,t} - \gamma_{2b,t}\right)\rho^{-t}, \tag{9.5}$$

where the term in parentheses is the number of units gained in year t, and H is the year when the enhancement begins ($H \geq 0$). To provide full compensation for the losses, $\gamma_{2a,t}$ must be set so that V_1 equals V_2. If the unit value terms are each constant over time, the equality is expressed as follows:

$$\sum_{t=0}^{T} \left(\gamma_{1b,t} - \gamma_{1a,t}\right)\rho^{-t} = \frac{v_1}{v_2} \sum_{t=H}^{T} \left(\gamma_{2a,t} - \gamma_{2b,t}\right)\rho^{-t}. \tag{9.6}$$

Obviously if the unit value terms are equal, the ratio v_1/v_2 can be ignored. In that case, the V terms give total discounted unit years.

[22]Federal statutes and most state laws require that the baseline recognize the natural change that can occur over time in the absence of the injury, but some states use the pre-injury condition as the baseline. In practice, a lack of data and modeling capability often results in the assumption of a static baseline (Viehman et al. 2009).

With habitat losses, as opposed to resource losses, service levels are lowered at the damaged spatial units. Analysis of habitat losses requires a different formulation (i.e., HEA), wherein the γ terms are not numbers of resource units as in REA, but rather service levels per spatial unit. In HEA, the present value of the habitat service losses, V_1, is

$$V_1 = q_1 \sum_{t=0}^{T} v_{1b,t} \left(\frac{\gamma_{1b,t}}{\gamma_{1b}} - \frac{\gamma_{1a,t}}{\gamma_{1b}} \right) \rho^{-t}, \tag{9.7}$$

where the damage occurs in year 0, q_1 is the number of spatial units injured, γ_{1b} is the average per-unit service level of the baseline in year 0, $\gamma_{1b,t}$ is the average per-unit service level of the baseline in year t, $\gamma_{1a,t}$ is the actual average per-unit service level in year t, $v_{1b,t}$ is the value of the average unit of the baseline in year t, and T and ρ^{-t} are as described above for REA. The ratio $\gamma_{1a,t}/\gamma_{1b}$ gives the proportion of the original (pre-injury) unit service level that is obtained in year t. The full term in parentheses gives the average per-unit service loss in year t as a proportion of the beginning service level.

As with REA, two simplifying assumptions are often applied. First, the baseline service level ($\gamma_{1b,t}$) is assumed to remain constant over time at the γ_{1b} level, such that $\gamma_{1b,t}/\gamma_{1b} = 1$ every year. Second, the average unit value ($v_{1b,t}$) is assumed to remain constant over the planning horizon. A further simplification is to—without actually measuring $\gamma_{1a,t,}$ or perhaps even γ_{1b}—directly estimate the ratio $\gamma_{1a,t}/\gamma_{1b}$, which is the proportion of the initial service level that is likely to be reached during each year of the recovery. With these three simplifying conditions, V_1 becomes

$$V_1 = q_1 v_{1b} \sum_{t=0}^{T} \left(1 - \theta_{1,t}\right) \rho^{-t}, \tag{9.8}$$

where $\theta_{1,t}$ is the directly estimated ratio $\gamma_{1a,t}/\gamma_{1b}$.

Knowing V_1, the present value of the lost services (Eqs. (9.7) or (9.8)), we now need to compute a similar level of service value for Option 2, the replacement. As with the damaged area, the present value of the enhancement at the replacement site, V_2, is a net of two service flows, as follows:

$$V_2 = q_2 \sum_{t=H}^{T} v_{2b,t} \left(\frac{\gamma_{2a,t}}{\gamma_{2b}} - \frac{\gamma_{2b,t}}{\gamma_{2b}} \right) \rho^{-t}, \tag{9.9}$$

where γ_{2b} is the average per-unit service level of the baseline in year 0, $\gamma_{2b,t}$ is the unit service level that will occur in year t without the enhancement, $\gamma_{2a,t}$ is the average per-unit service level that would occur in year t with the enhancement, $v_{2b,t}$ is the value of the average unit of the replacement baseline in year t, H represents the year when the enhancement begins ($H \geq 0$), and q_2 is the number of units of the replacement. As with $v_{1b,t}$, $v_{2b,t}$ remains constant into the future unless extant conditions (for example, about the scarcity of the injured service) suggest otherwise. To compensate for the service losses, V_1 must equal V_2. Given the unit values ($v_{1b,t}$ and $v_{2b,t}$), the unit physical levels of the services (the γ's), the time variables, the discount rate, and q_1, setting $V_1 = V_2$ reveals q_2.

As previously mentioned regarding Fig. 9.3, two basic possibilities exist for the relation of the actual (with enhancement) condition to the baseline (without enhancement) condition at the replacement site. In the absence of intervention, the site's resource service levels are rather stable at the initial level, but they can be improved with intervention (this case is shown in Fig. 9.3b). Alternatively, in the absence of intervention, the site's service levels are expected to diminish below the initial level (e.g., the site might be threatened with development), but the site can be protected so that the initial level of services is maintained (subject to some possible fluctuation; Chapman and Julius 2005). In the former case, $\gamma_{2b,t}$ remains relatively constant and $\gamma_{2a,t}$ increases above the beginning level, whereas in the latter case, $\gamma_{2a,t}$ remains relatively constant and $\gamma_{2b,t}$ drops below the beginning level. Of course, a combination ($\gamma_{2a,t}$ increases and $\gamma_{2b,t}$ decreases) could also occur. In all cases, the analysis would compute the difference in service levels between the enhanced and unenhanced situations. Given the former case, for example, and adopting similar simplifying conditions to those of Eq. (9.8), the following reduced equation for Option 2 is obtained:

$$V_2 = q_2 v_{2b} \sum_{t=H}^{T} \left(\theta_{2,t} - 1\right) \rho^{-t}, \tag{9.10}$$

where $\theta_{2,t}$ equals the directly estimated proportion of the initial baseline service level (γ_{2b}) that is likely to be reached in year t.

Adopting the simplified cases of Eqs. (9.8) and (9.10) and setting V_1 to equal V_2, q_2 is estimated as

$$q_2 = q_1 \frac{v_{1b}}{v_{2b}} \frac{\sum_{t=0}^{T} \left(1 - \theta_{1,t}\right) \rho^{-t}}{\sum_{t=H}^{T} \left(\theta_{2,t} - 1\right) \rho^{-t}}, \tag{9.11}$$

where the term v_{1b}/v_{2b} is the ratio of the two unit values.

With HEA and REA, the unit values typically reflect expert judgments of ecological value. If the unit values are assumed to be identical, the value ratio term of Eq. (9.6) or (9.11) drops out. If the values differ, the unit values express the quantitative relation between them. For example, if the initial baseline levels of per-hectare ecological services of a lost habitat are judged to be twice as important as those of the replacement habitat initial baseline level (perhaps, for example, the biological productivities of the sites differ), $v_{1b}/v_{2b} = 2$.[23]

[23]An important difference between VEA and HEA or REA is that changes over time in per-unit economic values of the injured and restored resources are generally overlooked in applications of HEA or REA (Dunford et al. 2004; Zafonte and Hampton 2007). Changes over time in economic

VEA can be thought of as a special case of either REA or HEA, where the unit values (e.g., v_1 and v_2 of Eq. (9.6), v_{1b} and v_{2b} of Eq. (9.11)) are estimated in monetary units. If the replacement resource or service is the same as the lost resource or service (e.g., Chinook salmon is replaced with other Chinook salmon), the replacement is said to be "in-kind," and the advantage of using VEA in place of REA or HEA is likely to be small. If the replacement resource or service is different from the lost resource or service (e.g., Chinook salmon is replaced with steelhead), the replacement is said to be "out-of-kind." With an out-of-kind replacement, forgoing an economic expression of the value terms should not be done lightly. As described in the regulations implementing the Oil Pollution Act of 1990, HEA is to be used only where it is judged that the proposed compensatory action provides services of "the same type and quality, and of comparable value" to those lost due to the injury (Chapman et al. 1998, p. 2).[24] If the equal-value requirement is judged to not have been met sufficiently, use of VEA could be warranted. In practice, however, scaling typically proceeds without measurement of economic values, and the v_1/v_2 term reflects scientific judgment of the relative contributions of the two service flows.[25]

The difference between in-kind and out-of-kind replacement is not carefully defined, and in fact, perfect comparability (i.e., achievement of $V_1 = V_2$) is unlikely. As the NOAA (1997) report explained, "Even when the proposed action provides the same type of natural resources and services, a variety of substitutions (in time, space, species, etc.) may be unavoidable. The result will be differences—in quality, in economic value, and in people who experience the service losses and those who experience the gains provided by the restoration actions" (p. 3-3). The differences could arise, for example, from a proposal to replace wild salmon with hatchery salmon or to replace a natural wetland with a constructed wetland. In the latter case, for example, even if a replacement wetland provides the same ability to assimilate wastes as the original wetland, the replacement wetland might not be located so as to provide humans with the services they originally enjoyed (e.g., one site may be upstream from a drinking water diversion point whereas the other is downstream, or one site may be recreationally valuable while the other is not due to lack of access). In the end, the degree to which comparability is accomplished in a certain compensatory restoration depends on many details of the situation at hand, including the

(Footnote 23 continued)

values can occur as population growth and economic development cause resource values to increase or as the damage causes a reduction in the supply of certain services leading to an increase in marginal value, which is likely to be followed with the recovery of the damaged resource plus the addition of the restoration resource by an increase in total supply and a resulting decrease in marginal value.

[24]This requirement is similar to the first of the three conditions for applying the RCM—that both options provide the same service(s).

[25]See Dunford et al. (2004), Flores and Thacher (2002), Jones and Pease (1997), Mazzotta et al. (1994), Randall (1997), Roach and Wade (2006), Unsworth and Bishop (1994), and Zafonte and Hampton (2007) for more on how HEA or REA compares with an economic valuation (VEA) approach.

options available and the professional judgment of the analysts about the ratio of the unit values (v_1/v_2).[26]

Where there are differences between the lost resources or services and those proposed for the replacement, if full economic valuation studies are not warranted, the specialists might consider using an inexpensive method, such as benefit transfer or a focus group or values jury (Brown et al. 1995) to gain a better understanding of public preferences, which can be used to inform the v_1/v_2 ratio.

9.2.2 The Metric

Application of EA is most straightforward when only one resource or service is affected. However, environmental damage typically affects multiple resources and/or services. Because quantifying the effects on the full set of resources and/or services would be difficult and costly, applications of EA typically select a single resource or service measure—a single metric, such as number of fish of a given species, herbaceous stem density per-unit area, acres of wetland, or gallons of water—to be used to represent the full set of affected resources and services (Chapman et al. 1998; Dunford et al. 2004; NOAA 2006). When a single metric is used, the γ measures of the above equations are for that metric. Because the analysis then focuses on the chosen metric (e.g., on how quickly the damaged site can return to providing the baseline level of the metric and on how easily the metric can be enhanced at a replacement site), its selection is critical to the success of the analysis. If the selected metric is not representative of the full set of damaged services, the required replacement activity could over or undercompensate for the losses (Milon and Dodge 2001). On the other hand, if the metric cannot be measured and predicted at a reasonable cost, it will also be ineffective and would not be appropriate for an EA.

When multiple resources or services have been affected, instead of selecting a single one to use as the metric, the metric can be a composite of a set of weighted individual measures. As explained by Viehman et al. (2009) regarding coral reefs, many approaches are possible, from relatively simple aggregations of two-dimensional coverages of individual plant species types (e.g., stony corals, octocorals, and sponges), to aggregations of frequency-size distributions for each species type, to aggregations of complex three-dimensional species-specific measures that account for the structural aspects of a coral reef.

[26]Comparability is not only constrained by lack of information about economic values. It can be optimistic to expect that ecological understanding has advanced to the point where ecological "equivalence" can be assessed with much greater efficacy than can economic value. The track record in re-creating wetlands, for example, is apparently not good (Barbier 2011, p. 1841-1843), and the metrics that have been used in HEA (e.g., percent of herbaceous cover) tend to be inadequate to assure re-creation of ecological structure and function.

Of course, the more complex the metric, the greater the amount of data (and perhaps modeling capability) required and the more challenging it will be to assess equivalency or the ratios of equivalencies. In selecting or designing the metric, the specialists must evaluate the trade-off between simplicity and realism. Further, when a composite metric is used, weights or conversion factors of some sort are needed to express the relative importance of each individual measure. No composite metric bypasses the need for weighting the different measures; thus, using a composite metric does not avoid some reliance on professional judgment. Rather, it substitutes the judgment involved in selecting a single service to represent the multiple affected services with the judgment required to weight the individual measures included in a composite metric.[27]

When two or more resources or services have been or will be damaged and no viable way is found to combine them into a single metric (whether based on a single measure or a composite of several measures), the option of employing more than one metric remains, which would require replacing more than one metric and conducting an EA for each unique component. In such cases, it could be useful to determine the amount of compensatory restoration necessary for each and then evaluate the combined restoration actions to identify potential overlap where the restoration project for one resource provides benefits for another damaged resource and adjust accordingly so as to end up with the most cost-effective restoration project.

9.2.3 Equivalency Analysis Scaling—an Example

The following hypothetical example, adapted from English et al. (2009), illustrates the key features of EA, most importantly the *scaling* of the enhancement activity at the replacement site so that the aggregate gains from the activity equal the aggregate losses caused by the injury. The loss in this example is caused by an offshore oil spill. Oil from the spill has killed 100 birds at sea and washed up on 10 ha of a salt marsh. It is estimated that the number of killed birds found at sea is only one-fifth of the number lost, resulting in an estimate of 500 birds lost due to the spill. The scaling analysis will express these two losses—the lost services of the marsh and the lost birds—with a single metric of habitat function so that a single replacement activity can be used. Converting the resource loss into a habitat loss allows use of HEA to assess both losses.

The metric selected for measuring the losses and the replacement gains is the service provided by 1 ha of fully functioning salt marsh over 1 year (i.e., a service-hectare-year, or SHY). Discounted SHYs are called DSHYs. A 50-year

[27]EA relies on professional judgment at many points, including selection of the metric. Because judgments can differ, involving several professionals, perhaps using a procedure such as Delphi might be advisable. When disagreement persists about a certain estimate, a resolution is sometimes reached by negotiation among the involved stakeholders (Ray 2009).

Table 9.4 Hypothetical example of per-hectare service losses at a salt marsh

Years since spill	Year-end service loss (%)	Midyear service loss (%)	Service loss (SHY)	Discount factor	Discounted service loss (DSHY)
0	75				
1	60	67.5	0.675	0.985	0.665
2	45	52.5	0.525	0.957	0.502
3	30	37.5	0.375	0.929	0.348
4	15	22.5	0.225	0.902	0.203
5	0	7.5	0.075	0.875	0.066
6	0	0.0	0.000	–	0.000
…					
50	0	0.0	0.000	–	0.000
Total			1.875		1.784

planning horizon is chosen ($T = 50$). Vegetative cover is selected as the indicator of the value of the services provided by a salt marsh. The condition of a nearby unaffected salt marsh is selected as the baseline, which is assumed to remain constant into the future. Comparison of the damaged salt marsh to the nearby unaffected site indicates that the oiled site has lost 75% of its vegetative cover. Based on prior experience, it is estimated that the oiled marsh will return to baseline condition in 5 years and that the improvement will be linear over time. Computation of the discounted service loss per hectare of salt marsh is shown in Table 9.4, assuming a 3% discount rate. The year midpoint is used as the measure of service loss in a given year. For example, the year 2 per-hectare service loss is estimated to be $(0.6 + 0.45)/2 = 0.525$ SHY, and the discounted service loss $0.526 \times 0.96 = 0.502$ DSHYs per hectare, or 5 DSHYs in total (given that 10 ha were oiled). In essence, a 52.5% service loss in year 2 on each of 10 hectares is assumed equal to a complete loss of services on 5.25 ha, which in present value terms is a 5.02-ha loss. Summing across years, the total discounted salt marsh service loss is 1.78 DSHYs per hectare or 17.8 DSHYs in total.[28] In terms of Eq. (9.8), $q_1 = 10$, v_{1b} is unspecified and essentially equal to 1, $T = 50$, $(1 - \theta_{1,t})$ = midyear service loss proportion (SHY of Table 9.4), and $r = 0.03$.

The bird population is assumed to recover in one year. To facilitate the computation of necessary compensatory actions, the bird loss must be converted into a loss of salt marsh DSHYs. The conversion, in this example, is based on the rate at which a salt marsh provides food for the affected birds. The marsh is estimated to provide 5,147 kg/ha/yr of food available to small sediment-dwelling invertebrates.

[28]Use of a 3% discount rate, which is common in REA, results in a discount factor of about 0.23 by Year 50 and 0.05 by Year 100. Because some species, such as slow-growing coral, can take even longer than that to fully recover (Viehman et al. 2009), the discounted service loss—even in the case where the planning horizon is ample—can be rather unreflective of the full extent of the physical loss.

Table 9.5 Hypothetical example of per-hectare service gains at a new salt marsh

Years since spill	Year-end service level (%)	Year-end service gain (%)	Mid-year service gain (%)	Discount factor	Discounted service gain (DSHY)
0	10.0	0.0			
1	10.0	0.0	0.0	0.99	0.00
2	10.0	0.0	0.0	0.96	0.00
3	10.0	0.0	0.0	0.93	0.00
4	15.3	5.3	2.7	0.90	0.02
5	20.7	10.7	8.0	0.88	0.07
6	26.0	16.0	13.3	0.85	0.11
…					
17	84.7	74.7	72.0	0.61	0.44
18	90.0	80.0	77.3	0.60	0.46
19	90.0	80.0	80.0	0.58	0.46
…					
49	90.0	80.0	80.0	0.24	0.19
50	90.0	80.0	80.0	0.23	0.19
Total					13.83

The invertebrates are eaten by fish, which are eaten by the birds. The trophic loss in the conversion from biomass to birds is estimated to be 99.6%. Thus, a fully functioning hectare of salt marsh is estimated to produce 5,147 kg × 0.004 = 20.6 bird kilograms in a year. Adult birds are estimated to weigh 2 kg each, so 1,000 kilograms of birds were lost (2 kg × 500 birds), which would require 1,000/20.6 = 48.6 SHYs of marsh-equivalent services to replace. Because the bird loss lasts only 1 year, discounting has little effect: 48.6 × 0.99 = 48.1 DSHYs. The total marsh-equivalent service loss is thus 17.8 + 48.1 = 65.9 DSHYs.

To offset the losses caused by the spill, new wetlands are to be created on nearby land. The site is estimated to currently provide 10% of the services, as measured by vegetative cover, of the damaged marsh before the spill. Excavation at the site to create the necessary land elevations and tidal flows is expected to allow the site to provide 90% of the services of the baseline condition in 15 years. The 90% service level is then predicted to remain constant over the rest of the 50-year planning horizon. Computation of the total discounted service gain per hectare is shown in Table 9.5, assuming that the excavation occurs in year 3, that the improvement in functioning is linear over time from year 3 to year 15, and a 3% discount rate. Because the baseline service level of the damaged marsh (γ_{1b}) was used as the baseline service level of the replacement marsh (γ_{2b}, i.e., the losses and gains are both specified in terms of a fully functioning salt marsh), the v_{1b}/v_{2b} term is assumed to equal 1. The total discounted gain is 13.8 service years per hectare. To provide gains equivalent to the loss, 65.9/13.8 = 4.8 ha of new salt marsh must be created. [29]

Obviously, application of EA requires professional judgments and technical estimates. In this example, the judgments involve selection of, among other things, the metric, the biological measure(s), the conversion approach, the replacement site, the discount rate, and the planning horizon. And the technical estimates include the extent of unseen impacts (additional birds killed), the rates at which conditions will improve at the two sites, the number of birds that a hectare of marsh can support, and the degree to which a created marsh can provide the services of a natural marsh. Results depend critically on these selections and estimates and can vary substantially depending on the choices made (Strange et al. 2002). Careful study of the damaged and undamaged sites, plus data on the effects of and recovery from prior events of similar characteristics, can help in making decisions about the required compensatory actions. But analysis is expensive, and the analyst must also consider that the cost of the analysis should be in keeping with value of the loss.

This example employs a simplistic measure of salt marsh losses—specifically, a single measure of total vegetative cover. In practice, an EA analysis is likely to divide the damaged area into different spatial (e.g., lightly, moderately, and heavily oiled) and perhaps also vertical (e.g., belowground soil condition in addition of aboveground vegetation) strata. The degree of loss of the different strata might then be determined by professional judgment or availability of scientific data to quantify injuries (such as biomass production for aboveground plants).

The example, as with applications of EA generally, does not attempt to account for uncertainty about the future success of compensatory restoration activities. Specialists provide their best estimates of future service gains, and the scale of the necessary activities is decided on that basis. However, a risk remains that things will not work out as well as estimated. Sensitivity analysis may help here to assess the impact of possible miscalculations. To lower the chance of falling short of expectations (i.e., to adopt a risk-averse position), the scale of the compensatory activities could be increased accordingly (Moilanen et al. 2009).

9.2.4 Steps in Applying Equivalency Analysis

To begin an EA, basic parameters of the analysis are chosen, most importantly the metric, the planning horizon (T), and the discount rate (r). These initial considerations and choices make up the first step of an EA effort (Table 9.6). The remaining steps are to (2) quantify the loss of ecological services caused by the natural resource injury; (3) settle on a restoration project and quantify the per-unit ecological benefits it will provide; (4) scale the restoration project in size to balance the

[29] The existence of two separate losses—the loss of services of the salt marsh and also the loss of birds at sea—requires adding a term to the numerator of Eq.(9.11).

Table 9.6 Steps of an equivalency analysis

Step 1	**Assess the losses and set basic parameters of the analysis**
	• Specify the resources and services affected
	• Decide if replacements of similar type, quality, and value are available (if they are not, consider the need for an economic valuation approach)
	• Decide on the metric to use in summarizing the gains and losses and scaling the replacement activity
	• Specify the planning horizon and the discount rate
Step 2	**Estimate the loss**
	• Determine the extent of the injury (e.g., q1 in HEA)
	• Decide on the specific resource or service to be measured (or otherwise characterized) to represent the metric[a]
	• Decide on the strata into which the affected area will be separated (if any)
	• Determine the baseline level of the metric, including (a) determine the current level of the metric, and (b) if the baseline is not considered static, estimate future levels of the metric in the absence of the damage
	• Settle on the primary restoration activity (or set of activities)
	• Determine the future levels of the metric given implementation of the primary restoration activity
Step 3	**Estimate the per-unit gain**
	• Select the replacement site(s) or resource
	• Decide on the beginning date (H) of the enhancement activities
	• Determine the unenhanced (baseline) level of the metric, including (a) determine the initial baseline level of the metric and (b) if the baseline is not considered static, estimate future levels of the metric
	• Settle on and schedule the enhancement activities
	• Determine the future levels of the metric given the enhancement activities
Step 4	**Scale the replacement activity to determine the quantity needed**
	• If the unit values are static, settle on the ratio of unit values; otherwise, include the unit values in the separate equations of V1 and V2
	• Solve for required quantities by setting V1 = V2
Step 5	**Determine the cost of the replacement activity for the entire time horizon**
	• List the specific actions, including and when and where they will occur
	• Determine if the responsible party will perform any of the activities
	• Estimate the cost of remaining activities

[a]A simple metric is assumed here. If a composite metric were used, several resources or services would be measured and weights would be needed for combining across those resources or services in order to convert them into a single metric

losses from the injury with the gains from the restoration; and (5) compute the cost of the restoration (Jones and Pease 1997; Zafonte and Hampton 2007).

Step 2 begins with assessing the full extent of the damage, including listing the potentially affected resources and services and characterizing the extent to which they have been or will be reduced by the damage. Next is to determine how the metric will be quantified. The metric may be represented by one or a combination of several measures, and it may involve breaking the injured site into several vertical or horizontal strata, each with its own measure of injury. This step, as with other steps involving quantification of levels of the metric, can involve use of an ecological or physical simulation model. In addition, the number of affected units of the metric (e.g., q_1 in HEA) is quantified. Next the baseline levels of that metric are quantified. With an ex post assessment, this is likely to require finding undamaged sites of similar character to the damaged site but for the damage. The current level of the baseline is quantified (γ_{1b}) and then, if the baseline levels are expected to change over time, the future levels are determined ($\gamma_{1b,t}$). Next the primary restoration activities are specified and, given implementation of those activities, the future levels of the metric given restoration are either estimated (thus determining $\gamma_{1a,t}$) or at least understood sufficiently to estimate the degree to which the restoration falls below the baseline level (resulting in estimates of $\theta_{1,t}$).

Step 3 begins with selecting the replacement site(s). This can require consideration and evaluation of several alternatives. As with the RCM, the least-cost acceptable alternative should be selected. Then, given the beginning date of the replacement activities (H), the initial baseline level at the replacement site is estimated (γ_{2b}). Selection of γ_{2b} is a critical step, because it affects the relation of v_{1b} to v_{2b}. If γ_{1b} is used as the baseline at the replication site (i.e., if $v_{1b} = v_{2b}$), then $v_{1b}/v_{2b} = 1$. However, if a separate baseline level of services is chosen, the ratio of the values (v_{1b}/v_{2b}) must be determined so that the gains and losses are comparable. If this level is expected to change over time, future levels are estimated throughout the planning horizon ($\gamma_{2b,t}$). Then the enhancement activities are finalized and either the effects of those activities on the metric over the planning horizon are estimated ($\gamma_{2a,t}$), or those effects are understood sufficiently to estimate the proportion of the baseline service levels that the activities will provide (giving $\theta_{2,t}$ in HEA).

In Step 4, the replacement quantity ($\gamma_{2a,t}$ in REA, q_2 in HEA) is determined by setting V_1 equal to V_2. In the simplest case, the unit values (v_{1b} and v_{2b}) are assumed to remain constant over time, and the ratio of those values (v_{1b}/v_{1a}) is used in Eq. (9.6) or (9.11). Of course, if the values are assumed to be equal, the value ratio term is unimportant. If the values are expected to vary over time, those values ($v_{1b,t}$ and $v_{2b,t}$) must have been used in their respective equations, and the required replacement quantities are then estimated by setting the resulting V_1 and V_2 equal to each other.

Finally, in Step 5 the cost of the replacement activity over the planning horizon is estimated. This will require listing all actions to be performed, along with the dates and locations of the actions. To assure that the cost is accurately estimated, it may be necessary to obtain estimates for accomplishing the actions from multiple sources.

9.2.5 Two Examples of Equivalency Analysis Application

EA has been used in a wide variety of situations, and in response to several different statutes requiring compensation for losses. Two examples of EA used in natural resource damage assessments in the United States are described here. The first example is of a resource loss, where the compensatory actions improve resource numbers beyond their static baseline, and the second example is of a damaged habitat where the compensatory actions avoid a future reduction in the quality of the replacement area. For additional examples, see English et al., (2009), who summarized applications for four kinds of organisms (fish, birds, reptiles, and mammals) and eight different kinds of habitats (coastal marsh, sedimentary benthic habitat, water column, oyster reef, mangroves, seagrass, coral reef, and kelp forest).

9.2.5.1 The Blackbird Mine

Mining at the Blackbird site beginning in the 1890s left behind tailings, an open pit, underground tunnels, and other disturbances that continued to pollute streams in the Panther Creek drainage, a tributary of the Salmon River in Idaho, long after mining ended in the late 1960s (Chapman et al. 1998). In 1992, the state of Idaho initiated a natural resource damage assessment, which was joined by two federal agencies: the National Oceanic and Atmospheric Administration and the U.S. Forest Service. Studies found releases of hazardous substances, primarily cobalt and copper. Comparison of conditions in the affected streams with conditions in nearby unaffected streams showed that the lowering of water quality caused by the mining remains had significantly lowered numbers of cutthroat, bull, brook, and rainbow trout and had eliminated Chinook salmon runs.

Injured resources included surface water, streambed fauna, and resident and anadromous fish. Service losses included lost recreational and subsistence fishing opportunities and diminished quality of other kinds of outdoor recreation, such as hiking and camping. The trustees selected the REA method, using naturally spawning spring/summer Chinook salmon as the metric for scaling primary and compensatory restoration. The selection was based on the assumption of a high correlation of salmon vitality to overall ecosystem health. The selection was all the more salient because Salmon River Chinook salmon had been listed as a threatened species under the Endangered Species Act. Taking into account other limiting factors on Chinook salmon numbers (e.g., downstream dams and fishing pressure), it was estimated that Panther Creek and its tributaries, in the absence of the contamination, would support annual runs of 200 adult Chinook salmon. The baseline was set at a constant level of 200 spawning adults annually. Several restoration actions were specified to bring the area up to the 200 Chinook salmon level, including cleanup at the mine site to improve stream water quality, reintroduction of naturally spawning salmon into Panther Creek, and various activities to improve the smolt survival rate, including channel meander reconstruction and riparian corridor fencing. To estimate the interim losses, damage was counted as beginning in 1980, the year of passage of the Comprehensive Environmental Resource, Compensation, and Liability Act of 1980.[30] Primary restoration was predicted to allow Chinook salmon to begin using the site in 2005 and to reach the baseline of 200 in 2021. Thus, interim losses would accrue over the period 1980–2020. Activities chosen to compensate for the expected interim losses included efforts to raise the number of Chinook salmon in Panther Creek above 200 beginning in 2021 and fencing of riparian areas outside of the Panther Creek drainage to improve salmon habitat there. The cost of those activities constituted the compensatory part of the assessed damage payment.

In terms of the model presented earlier and using the resource loss version of the model, V_1 of Eq. (9.4) was computed assuming a constant baseline (γ_{1b}) of 200 adult Chinook salmon and an initial complete loss of the salmon ($\gamma_{1a,t=0} = 0$). Damage began in 1980 ($t = 0$). Actual levels of Chinook salmon in Panther Creek were 0 until 2005 ($\gamma_{1a,t} = 0$ for $t < 26$), and rose to 200 in 2021 ($\gamma_{1a,t} = 200$ at $t = 41$), with $\gamma_{1a,t}$ for $t > 25$ determined by a Chinook population model. Compensatory restoration was in two parts. In Panther Creek it began in 2021 ($H = 41$) when Chinook salmon numbers were predicted to rise above 200 due to riparian corridor fencing and other improvements. Outside of Panther Creek the enhancement effors may have begun earlier. Both within and outside of Panther Creek drainage, the enhancements were assumed to improve conditions above a static baseline. The total compensation effort was scaled so that the discounted salmon years gained from the two enhancement efforts equaled the discounted salmon years lost, using a 3% discount rate. Because the loss and gains were both in terms of wild Chinook salmon in and around Panther Creek, this analysis was considered in-kind and the unit values (v_1 and v_2) were assumed equal and, therefore, of no consequence in the scaling equation.[31] This determination was supported in part by the judgment that the total number of salmon at issue, roughly 200, was too small to affect the unit value of salmon in the larger Snake River drainage.

One might say that the trustees, in facing the trade-off between simplicity and realism, overly emphasized simplicity, given their decision to focus on a single species, Chinook salmon. However, accurately modeling the response of Chinook salmon to changes in water quality and other habitat measures is itself extremely difficult and fraught with uncertainties; evaluating additional species to create a composite metric would have compounded the difficulty, with no assurance that the result would have been better than relying on Chinook salmon numbers as the indicator of overall ecosystem health. Such are the complex choices encountered in any real-life application of EA.

---

[30]For legal reasons it is common in U.S. damage assessment to ignore damages that occurred before passage of the enabling legislation.

[31]The reintroduced Chinook salmon originated from wild salmon donor stocks from an adjacent drainage (rather than being common hatchery salmon).

9.2.5.2 The Storrie Fire

The 21,000-ha Storrie forest fire on national forest land in California in 2000 was caused by a spark from a repair operation on a railroad right-of-way. In addition to recovering the cost of suppressing the fire, the value of the timber that could have been harvested had the fire not occurred, and the cost of restoration (replanting), the trustee sought to recover the value of additional injuries associated with the interim loss of scenery, wildlife habitat, soil quality, water quality, and recreational use (Kimball 2009). The trustee claimed that the fire burned more than 1,600 acres of spotted owl habitat, 12,000 acres of carnivore habitat, and 9,000 acres of old growth forests, impacting several important bird species (bald eagles, goshawks, and pine martens) as well as amphibians and fish in streams draining the area, and lowering scenic beauty along highways, trails, and a scenic byway.

The trustee used HEA to estimate the required compensation for the interim losses.[32] The metric adopted for measuring those losses was the quadratic mean diameter (QMD) of the trunks of the trees lost per acre,[33] and the replacement activity chosen was fuels management on nearby forestland (the replacement area). Fuels management was predicted to lessen the chance of future forest fires in the replacement area. Thus, the compensation requested was the cost of sufficient fuels management to avoid future loss of trees, again measured in terms of QMD, equivalent to the interim losses in QMD caused by the fire.

The HEA application accounted for different overstory types and densities (i.e., different spatial strata), but for simplicity here this description assumes only one overstory situation. The present value of the losses, V_1, was computed as in Eq. (9.7) with the baseline set equal to and held constant at the QMD of the average acre prior to the fire (such that $\gamma_{1b,t}/\gamma_{1b} = 1$ in all years t). The QMD in the burned area as the forest recovered ($\gamma_{1a,t}$) was estimated using a vegetation simulation model, and $\gamma_{1a,t}/\gamma_{1b}$ was the ratio of predicted QMD in year t to the baseline level. Other parameters were as follows: q_1 was set at the number of burned acres; T was set at 2101, by which time stand recovery was predicted to be mostly complete; r was 0.03; and $v_{1b,t}$ was ignored (as was $v_{2b,t}$) because a lost QMD in the burned area was considered of equivalent quality and value to a gained QMD in the replacement area. The sum, over years 2001–2101, of the discounted annual estimates of proportion of QMD lost on the representative acre, multiplied by the size in acres of the burned area, gave the present value of the total number of lost acre-years (i.e., V_1).

The per-acre present value of gains at the replacement site, V_2/q_2, was then estimated as in Eq. (9.9), where the baseline QMD level ($\gamma_{2b,t}$) assumed no fuel treatments to lower the probability of a catastrophic wildfire in the replacement forest stand, and the enhancement QMD level ($\gamma_{2a,t}$) reflected a reduction in the

---

[32]This summary is based largely on personal communication with Robert Unsworth (June 2013) of Industrial Economics Inc., who prepared an Expert Report for the plaintiff in the Storrie fire case.

[33]The quadratic mean weights larger trees more heavily than would a simple mean, which was understood to more accurately reflect the services that flow from trees of different sizes. Tree diameter is measured at breast height (about 1.5 m from the ground).

likelihood of wildfire due to the proposed fuel treatments. Specifically, V_2/q_2 was computed assuming that without fuel treatments, the probability of a forest fire would be 0.0053 each year over the 100-year planning horizon, conditional on the area not having already burned, but that with fuel treatment, the probability would be 0 for 15 years and 0.0053 thereafter (again conditional on no prior burn).[34] For both cases, the discounted expected service levels over the 100 years were computed and then summed. The difference between the two sums yielded an estimate of the discounted expected per-acre benefit of the fuel treatment. Furthermore, it was assumed that treating 1 acre would protect an average of 3.33 acres from catastrophic wildfire (assuming that the treatments were well placed within the larger area), thus lowering the acres needing treatment. The number of acres to be treated, q_2, was chosen to equate V_1 with V_2. Multiplying q_2 by the per-acre cost of treatment yielded the total cost. However, fuel treatment was assumed to provide "collateral" benefits—mostly in terms of avoided future fire suppression and reforestation costs—the present value of which was subtracted from the total cost, thereby further lowering the number of acres that would actually be treated.

The Storrie fire analysis relied on a forest vegetation model for predicting QMD levels 100 years into the future and, as indicated, on a host of assumptions about stand conditions, costs, and collateral benefits. Some of the assumptions tended to produce a conservative estimate of the compensation needed, such as that the baseline QMD in V_1 did not increase over time as the trees continued to grow. Other assumptions tended to increase the required compensation, such as that noncatastrophic (indeed, beneficial) fires would not occur in the replacement area, thus reducing the benefit from fuel treatments. Further, no benefits were attributed to the Storrie fire, although the fire undoubtedly lowered future fire suppression and related costs. Attributing no benefits to the damaging event would be quite reasonable if the event were a totally unnatural occurrence, such as an oil spill or mine drainage, but forest fires occur naturally, and the only unnatural aspects of the Storrie fire were the time when it occurred and the fuels present at that time (for example, the fuels could have been in excess of natural conditions due to past fire suppression efforts and to construction of flammable structures such as houses). Thus, the Storrie fire example illustrates both the complexities (and many assumptions) involved in matching environmental gains and losses and the unique challenges introduced when the damaging event could have occurred naturally (such as when forest fire is ignited by lightening).

9.3 Summary and Closing Thoughts

The replacement cost method and equivalency analysis are both substitution methods in the general sense that they focus on alternative ways to provide a given resource or service. However, the role of the substitute differs between the methods.

[34]The 0.0053 estimate was based on a 102-year fire history during which 27,531 acres burned out of a total of 50,624 acres. It was further assumed that once the area had burned, it would not burn again.

With the RCM, the substitute (Option 2) will certainly be provided if the service at issue (Option 1) is not, but with EA, Option 1 has been or will be lost. Thus, technically Option 2 in EA is a replacement, not a substitute, for Option 1.

Use of the RCM can provide a measure of the gross benefit (economic value) of the services provided by a proposed project or protected resource. The method equates that benefit measure with the cost of providing a substitute, but only if the following three conditions are satisfied: (1) the substitute provides equivalent services, (2) the substitute is the least-cost alternative way to provide the services, and (3) it is clear that the services provided by the substitute would be demanded at the specified cost if those services were not provided by the project or resource being studied. The RCM has been used in two basic situations: to value a new public investment and to value protection of an existing natural resource. Generally, the confidence that the third condition has been met, and thus the confidence that the estimated cost can be accepted as a measure of benefit, is greater when valuing a new public investment than when valuing protection of an existing resource. However, regardless of confidence in the benefit estimate and even if an application clearly fails to provide a valid estimate of the benefit, an application can provide a cost estimate indicating whether or not the proposed project or resource protection would enhance social well-being under the existing institutional framework.

EA does not aspire to provide an economic measure of benefit. Rather, EA is used to quantify the payment needed to create or protect an environment that will offer services equivalent to services that have been or will be lost. In its most commonly used variants (HEA and REA), EA focuses on resources or ecological services that are lost and subsequently gained and does not require economic valuation. In the course of estimating a replacement cost, EA computes V_1, the biophysical change in resource or service levels, with no assurance that the economic value of the losses, were it to be estimated, exceeds the replacement cost.

The first of the three conditions for use of the RCM (that the two options provide equivalent services) is critical to EA as well. With the RCM, the substitute must provide services of equal value to the services provided by the project or resource at issue, whereas with EA, the replacement must provide equivalent ecological resources or services to those lost. Regardless of method, while simple in concept, meeting this condition is difficult in practice because of the many differences that can arise. To the extent that such differences are not accounted for in the specification of the substitute, the RCM fails to provide a valid measure of benefit and both methods may fail to provide sufficient information to allow a decision about the proposed project, resource protection, or service restoration at hand.

The other two conditions for use of the RCM also have corresponding relevance to EA. Regarding the second condition, just as the alternative (Option 2) must be the least-cost way to provide equivalent services with the RCM, the replacement in EA must be provided at least cost. The third condition for use of the RCM ($b_2 \geq c_2$) would in EA require that the replacement (Option 2) be desired if it were to cost as much as estimated. Because EA focuses on replacing ecological resources or services, this condition is not in fact imposed. Because of this lack of concern about the net economic value of the replacement, one may suspect that the third condition is satisfied in EA with roughly the same (low) level of confidence as in a typical application of the RCM when the RCM is used to value protection of an existing resource condition. That is, in many cases the only support for accepting $b_2 \geq c_2$ in the RCM or EA is the existence of a law implying that in general the benefits exceed the costs. Of course, with an EA-based damage assessment or a protection-based RCM, an estimate of b_2, and therefore of b_1, is of no practical importance if a mandate is present because the replacement must be provided and the task before the public agency is essentially one of choosing the less expensive way to replace the lost services. In such situations, the role of economic values is most important in informing the legislative and rule-making processes that establish such mandates rather than in implementing the mandate.

It is clear that the use of substitution methods relies on a good deal of professional skill and judgment. The ease with which the RCM, for example, can be misused has engendered important cautions about its use (e.g., Herfindahl and Kneese 1974). However, judgment calls are necessary and the opportunity for misuse exists with all economic valuation methods and, indeed, with empirical studies generally. Regardless of the method, the analyst must aspire to produce a credible result and must recognize when use of the method is inappropriate or when the available data are inadequate to support its use. Following the guidance provided here can enhance the credibility of the empirical applications of substitution methods.

Many ecosystem services—erosion control, stream water quality maintenance, and pollination to name only three—may best be valued using a supply-side valuation method (Brown et al. 2007). As demand for environmental services increases and those services consequently become more valuable, and as the role of ecosystem services in supporting human societies becomes more completely understood, the call for establishing ecosystem service values in economic terms will grow, and substitution methods, along with other supply-side methods, are likely to be employed.

Acknowledgments This chapter has benefited greatly from reviews by David Chapman, Ikuho Kochi, Mark Buckley, and John Loomis.

References

Adger, W. N., Kelly, P. M. & Tri, N. H. (1997). Valuing the products and services of mangrove restoration. The Commonwealth Forestry Review, 76, 198-202.

Barbier, E. B. (2011). Coastal wetland restoration and the Deepwater Horizon oil spill. Vanderbilt Law Review, 64, 1821-1849.

Brander, L. M., Wagtendonk, A. J., Hussain, S. S., McVittie, A., Verburg, P. H., de Groot, R. S. & van der Ploeg, S. (2012). Ecosystem service values for mangroves in Southeast Asia: A meta-analysis and value transfer application. Ecosystem Services, 1, 62-69.

Brown, T. C., Bergstrom, J. C. & Loomis, J. B. (2007). Defining, valuing, and providing ecosystem goods and services. Natural Resources Journal, 47, 329-376.

Brown, T. C., Peterson, G. L. & Tonn, B. E. (1995). The values jury to aid natural resource decisions. Land Economics, 71, 250-260.

Carson, R. T., Mitchell, R. C., Hanemann, M., Kopp, R. J., Presser, S. & Ruud, P. A. (2003). Contingent valuation and lost passive use: Damages from the Exxon Valdez oil spill. Environmental and Resource Economics, 25, 257-286.

Chapman, D., Iadanza, N. & Penn, T. (1998). Calculating resource compensation: An application of the service-to-service approach to the Blackbird Mine hazardous waste site. National Oceanic and Atmospheric Administration, Damage Assessment and Restoration Program. Technical paper 97-1. Washington, DC: NOAA.

Chapman, D. J. & Julius, B. E. (2005). The use of preventative projects as compensatory restoration. Journal of Coastal Research, Special Issue 40, 120-131.

Chatterjee, A. (2013). Annual crop residue production and nutrient replacement costs for bioenergy feedstock production in United States. Agronomy Journal, 105, 685-692.

Clarke, K. D. & Bradford, M. J. (2014). A review of equivalency in offsetting policies. Canadian Science Advisory Secretariat. Research Document 2014/109. St. Johns, Newfoundland, Canada: Fisheries and Oceans Canada.

Committee to Review the New York City Watershed Management Strategy. (2000). Watershed management for potable water supply: Assessing the New York City strategy. Washington, DC: National Research Council.

de Groot, R., Brander, L., van der Ploeg, S., Costanza, R., Bernard, F., Braat, L., … van Beukering, P. (2012). Global estimates of the value of ecosystems and their services in monetary units. Ecosystem Services, 1, 50-61.

Dunford, R. W., Ginn, T. C. & Desvousges, W. H. (2004). The use of habitat equivalency analysis in natural resourc damage assessments. Ecological Economics, 48, 49-70.

Eckstein, O. (1958). Water-resource development: The economics of project evaluation. Cambridge, MA: Harvard University Press.

English, E. P., Peterson, C. H. & Voss, C. M. (2009). Ecology and economics of compensatory restoration. Durham, NH: NOAA Coastal Response Research Center.

EPA (U.S. Environmental Protection Agency). (2009). Valuing the protection of ecological systems and services: A report of the EPA science advisory board. Washington, DC: EPA.

Flores, N. E. & Thacher, J. (2002). Money, who needs it? Natural resource damage assessment. Contemporary Economic Policy, 20, 171-178.

Freeman, A. M. (2003). The measurement of environmental and resource values: Theory and methods (2nd ed.). Washington, DC: Resources for the Future.

Goldberg, V. P. (1994). Recovery for economic loss following the Exxon Valdez oil spill. Journal of Legal Studies, 23, 1-39.

Griffin, R. C. (2012). The origins and ideals of water resource economics in the United States. Annual Review of Resource Economics, 4, 353-377.

Hanson, D. A., Britney, E. M., Earle, C. J. & Stewart, T. G. (2013). Adapting habitat equivalency analysis (HEA) to assess environmental loss and compensatory restoration following severe forest fires. Forest Ecology and Management, 294, 166-177.

Haveman, R. H. (1972). The economic performance of public investments: An expost evaluation of water resource investments. Washington, DC: Resources for the Future.

Hein, L. (2011). Economic benefits generated by protected areas: The case of the Hoge Veluwe Forest, the Netherlands. Ecology and Society, 16(2), 1-19.

Herfindahl, O. C. & Kneese, A. V. (1974). Economic theory of natural resources. Columbus, OH: Merrill.

Jackson, S., Finn, M. & Scheepers, K. (2014). The use of replacement cost method to assess and manage the impacts of water resource development on Australian indigenous customary economies. Journal of Environmental Management, 135, 100-109.

Jones, C. A. & Pease, K. A. (1997). Restoration-based compensation measures in natural resource liability statutes. Contemporary Economic Policy, 15(4), 111-122.

Kimball, S. S. (2009). Forest fire damages in transition. The Federal Lawyer, 56(7), 38-46.

King, D. M. & Adler, K. J. (1991). Scientifically defensible compensation ratios for wetland mitigation. Washington, DC: U.S. Environmental Protection Agency.

Ledoux, L. & Turner, R. K. (2002). Valuing ocean and coastal resources: A review of practical examples and issues for further action. Ocean & Coastal Management, 45, 583-616.

Mazzotta, M. J., Opaluch, J. J. & Grigalunas, T. A. (1994). Natural resource damage assessment: The role of resource restoration. Natural Resources Journal, 34, 153-178.

Milon, J. W. & Dodge, R. E. (2001). Applying habitat equivalency analysis for coral reef damage assessment and restoration. Bulletin of Marine Science, 69, 975-988.

Moilanen, A., van Teeffelen, A. J. A., Ben-Haim, Y. & Ferrier, S. (2009). How much compensation is enough? A framework for incorporating uncertainty and time discounting when calculating offset rations for impacted habitat. Restoration Ecology, 17, 470-478.

National Research Council. (2005). Valuing ecosystem services: Toward better environmental decision-making. Washington, DC: National Academies Press.

NOAA (National Oceanic and Atmospheric Administration). (1997). Scaling compensatory restoration actions: Guidance document for natural resource damage assessment under the Oil Pollution Act of 1990. Silver Spring, MD: NOAA.

NOAA (National Oceanic and Atmospheric Administration). (2006). Habitat equivalency analysis: An overview. NOAA Damage Assessment and Restoration Program. Washington, DC: NOAA.

Point, P. (1994). The value of non-market natural assets as production factor. In R. Pethig (Ed.), Valuing the environment: Methodological and measurement issues (pp. 23-57). Boston, MA: Kluwer.

Quigley, J. T. & Harper, D. J. (2006). Compliance with Canada's Fisheries Act: A field audit of habitat compensation projects. Environmental Management, 37, 336-350.

Randall, A. (1997). Whose losses count? Examining some claims about aggregation rules for natural resource damages. Contemporary Economic Policy, 15(4), 88-97.

Ray, G. L. (2009). Application of habitat equivalency analysis to USACE projects. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Roach, B. & Wade, W. W. (2006). Policy evaluation of natural resource injuries using habitat equivalency analysis. Ecological Economics, 58, 421-433.

Shabman, L. A. & Batie, S. S. (1978). Economic value of natural coastal wetlands: A critique. Coastal Zone Management Journal, 4(3), 231-258.

Steiner, P. O. (1965). The role of alternative cost in project design and selection. Quarterly Journal of Economics, 79, 417-430.

Strange, E., Galbraith, H., Bickel, S., Mills, D., Beltman, D. & Lipton, J. (2002). Determining ecological equivalence in service-to-service scaling of salt marsh restoration. Environmental Management, 29, 290-300.

Unsworth, R. E. & Bishop, R. C. (1994). Assessing natural resource damages using environmental annuities. Ecological Economics, 11, 35-41.

Viehman, S., Thur, S. M. & Piniak, G. A. (2009). Coral reef metrics and habitat equivalency analysis. Ocean & Coastal Management, 52, 181-188.

Young, R. A. (2005). Determining the economic value of water: Concepts and methods. Washington, DC: Resources for the Future.

Young, R. A. & Gray, S. L. (1972). Economic value of water: Concepts and empirical estimates. Technical Report to the National Water Commission. Springfield, VA: National Technical Information Service.

Zafonte, M. & Hampton, S. (2007). Exploring welfare implications of resource equivalency analysis in natural resource damage assessments. Ecological Economics, 61, 134-145.

Chapter 10
Experimental Methods in Valuation

Craig E. Landry

Abstract This chapter discusses the role of behavioral experiments in evaluation of individual economic values. Principles of experimental design play a role in application and assessment of non-market valuation methods. Experiments can be employed to assess the formation of preferences and the role of personal characteristics, social factors, and economic constraints on economic values. Experiments can be used to test the efficacy of nonmarket valuation methods and to study the effect of the valuation task, information, and context on valuation responses. We discuss these issues in turn, incorporating pertinent literature, to provide a review and synthesis of experimental methods in valuation.

Keywords Nonmarket valuation · Experiments · Preference formation · Measurement · Valuation efficacy · Contextual effects

Experimental methods are the cornerstone of scientific inquiry. First documented in the natural sciences in the early seventeenth century (Levitt and List 2007), methods of experimentation involve systematic procedures to evaluate fundamental propositions by controlling external factors in an attempt to isolate cause-and-effect relationships. Experiments in social sciences provide a powerful tool to address common problems in traditional empirical analysis, in particular, the lack of counterfactuals. Individuals, groups, and firms are never observed simultaneously in two states of the world. Randomization of subjects to control and treatment conditions can provide appropriate counterfactuals so that average treatment effects can be identified.

C.E. Landry (✉)
University of Georgia, Athens, GA, USA
e-mail: clandry@uga.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_10

In exploring price discovery via tâtonnement (iterative, trial-and-error process by which equilibria can be established) and building on the work of Professor Edward Chamberlain,[1] Nobel Laureate Smith (1964) was a pioneer in applying laboratory experiments to economics. Early work in experimental economics included analysis of strategic interactions in game settings (e.g., the classical prisoner's dilemma) and investigation of individual choices, such as behaviors predicted by expected utility theory (Davis and Holt 1993). By the late 1990s, experimental economics had become widely practiced, and experimentation was gaining general acceptance as a valid approach for drawing inference, though controversies remained regarding the external validity (or the generalizability) of lab results, the role of context and social norms, and the interplay between laboratory and field experiments (Levitt and List 2007).

Economic experiments entail observation of human behavior in a way that facilitates comparisons, allowing a statistical test to be performed and inferences to be drawn. Comparisons can be made with a predicted behavior, such as a theoretical expectation, or another observed behavior. In the classic economic experiment, groups of subjects are randomly placed in situations that differ in some carefully selected way, allowing inference about the effect of the difference between the situations. It is also common to expose individual subjects to multiple stimuli and compare responses to the various stimuli within a single setting. The observed behaviors could be in response to a controlled manipulation or could occur due to natural variation in some field setting.

In this chapter, we explore the relevance of experimentation to economic valuation, with a focus on nonmarket values. As with other economic experiments, nonmarket valuation experiments can take place in the lab or the field; indeed, field applications of nonmarket valuation methods can become experiments as well, when carefully juxtaposed conditions are implemented.[2] In addition, experimentation is relevant to nonmarket valuation in that the principles of experimental design can play a key role in applying nonmarket valuation methods.

Experiments have been used to study three basic topics relevant to nonmarket valuation. First, experiments are used to understand the formation of preferences and values, as well as the effect of personal characteristics or circumstances, such as perceived social norms, time or income constraints, and risk attitudes on those values. Second, experiments are used to test the efficacy of nonmarket valuation methods. For example, the hypothetical payment obtained using a stated preference method can be compared with actual payment in an experiment that places subjects in situations that differ only in whether hypothetical or actual payment is elicited. Third, experiments are used to study the effect of the valuation task, information, and context on response, behavior, and value. For example, the effects of such

[1]Chamberlain taught economics at Harvard from the 1930s to the 1960s, where he conducted what might have been the first economic experiments, initially as classroom exercises, to demonstrate the formation of prices. See Roth (1995) for more about the history of experimental economics.

[2]Examples include the use of field experiments to value water filters in rural Ghana (Berry et al. 2015) and bednets for protection from insects in Uganda (Hoffman 2009; Hoffman et al. 2009).

elements as scope of the good or policy being valued, response format, and scenario complexity have been studied.

Most nonmarket valuation experimentation has focused on stated preference methods. Revealed preference analyses do not typically make use of experimental methods, but in some instances, "quasi-randomization" due to exogenous events or policies can provide a natural experiment (Meyer 1995). Such natural experiments compare outcomes across treatment and control groups that are not randomly assigned by the researcher. Rather there is a transparent and exogenous source of variation that determines treatment assignment. For example, Hallstrom and Smith (2005), Carbone et al. (2006), and Bin and Landry (2013) used the occurrence of hurricanes to examine the impact of updated flood risk information on residential property values, with properties in designated flood zones serving as treatment groups and properties outside of flood zones serving as a control group. As in all empirical work, threats to internal, external, and construct validity abound. A clear understanding of sources of variation and trends in data, the possibility of omitted or endogenous variables, the likelihood and magnitude of measurement errors, and the prospect of selection effects are critical to constructing effective research designs using natural experiments (Meyer 1995).

The following sections (1) review the elements of experimental design, (2) consider the role of experimental design in applying nonmarket valuation methods, and (3) demonstrate the application and utility of experimentation via a review of selected experiments used to examine important issues in nonmarket valuation research.

10.1 Elements of Experimental Design

This section provides a succinct overview of experimental design, covering terminology and such topics as control and replicability, orthogonality, and subject payments. Readers seeking a more substantive background are encouraged to consult one of numerous tomes on experimental economics (e.g., Davis and Holt 1993; Kagel and Roth 1995). Particular topics likely to be of most interest for nonmarket valuation include auctions (Davis and Holt, Chaps. 3 and 5; Kagel 1995), public goods and voting (Davis and Holt, Chap. 6; Ledyard 1995), externalities (Davis and Holt, Chap. 6), and decision-making under risk (Davis and Holt, Chaps. 2 and 8; Camerer 1995). Common among all of these topics are two fundamental concepts of experimental design: control and replicability.

10.1.1 Control and Replicability

Control refers to the ability to manipulate and maintain conditions under which choices are made so that observed behavior can be used to test theoretical propositions and infer treatment effects. Effective control neutralizes any factors that would corrupt inference of treatment effects, such as incommensurability of subject pools or collinearity between treatment and other independent variables.

Replicability relates to procedural regularity, in which all details of the research protocol are reasonably justified and clearly described. Replicability is not unique to experimental methods; it applies to any field in which primary data are collected (and in many cases where secondary data are analyzed). The standards for replication, however, are strict within the discipline of experimental economics. Procedures and methods—for subject recruitment, subject pool design (e.g., nature and experience levels of subjects), provision of background and contextual information, subject instructions, types of incentives provided, and session duration, among other factors—have important implications for the validity and interpretation of results.[3] Clear documentation of all methods and procedures permits evaluation of the research, replication for verification purposes, and further refinement and expansion of the experimental hypotheses and testing procedures.

To date, deception of subjects has been universally proscribed in experimental economics, and the use of deception typically results in harsh criticism at academic conferences and rejection of manuscripts submitted to archival economic journals. The rationale behind this stringent standpoint is simple: economic experiments are designed to evaluate how subjects make decisions, some of which are highly complex. The general view in the profession has been that subjects need to believe that protocols are as specified; otherwise researchers will lack confidence that patterns in observed choices reflect underlying experimental conditions. Rather, they could reflect subjects' intuitive construction of the protocol (which could reflect perceived deceitful intent). Essentially, deception can lead to loss of control.

10.1.2 Terminology and Basic Experimental Designs

Davis and Holt (1993) adopted the following terminology, which is fairly standard in experimental economics

cell—A collection of sessions with the same treatment conditions.
cohort—A collection of subjects that participates in a session.
experiment—A collection of sessions in one or more cells, the composition of which reflects the intent of the study.
session—A sequence of periods involving individual or group decision tasks that utilize a particular group of subjects (usually within a single day).
treatment—A distinct environment or configuration of control variables (e.g., information provided, decision rules, experience level, subject type) actualized as a distinct experimental protocol.

[3] See Davis and Holt (1993, Chap. 1), for more details.

An experimental design is a specification of sessions in one or more cells to assess propositions related to study objectives. Treatments with baseline conditions are often referred to as control cells or simply "controls." The other cells are referred to as "treatment cells," although the term "treatment" is sometimes used in a general way to indicate all cells, whether they are control or treatment cells.

In estimating treatment effects, the nature of the inference that may be drawn depends on whether a "between" or "within" design is used. A between design compares decisions and outcomes between pools of subjects exposed to separate treatments. Essentially, between designs employ comparisons across cells. It is possible, however, to use a between design in a single session, varying some element of the experimental protocol across subjects (typically private information of individual subjects).

In contrast, within designs compare decisions and outcomes among a unique set of subjects in the same (and possibly across) session(s). This is made possible by exposing all subjects in the session(s) to the same set of treatments. A within treatment can be thought of as a special case of block design in which the experimenter blocks on a single subject, implementing treatment and control or multiple treatments for that single subject (List et al. 2011). An advantage of this approach is that it can greatly reduce the error variance, increasing experimental power. Nonetheless, some researchers cast aspersion on this design because it permits correlation of treatment effects with other dynamic phenomena (such as learning, subject fatigue, and wealth effects). When exposing subjects to multiple conditions or stimuli within treatments, it is imperative to vary the order of exposure across subjects or sessions in order to test for, analyze, and control order effects.[4] (By varying order, in essence, the within design becomes a between design, where order of treatments is a design factor.)

10.1.3 Orthogonality, Randomization, Sample, and Cell Sizes

In designing experiments, it is extremely important that orthogonality be maintained among control and treatment cells. In economics and statistics, orthogonality refers to independence among covariates (and unobservable factors) that affect a

[4]A robust experimental design that employs within-subject comparisons will also provide for between-subject comparisons with a changing of the ordering of stimuli. For example, consider an experiment employing two sets of stimuli, A and B, presented in treatments that switch the order of the stimuli sets. Thus we have the following treatments: 1.A-B stimuli ordering and 2.B-A stimuli ordering, which allow for: (i) within comparisons—1.A-B and 2.B-A; (ii) testing of order effects—1.A versus 2.A, and 1.B versus 2.B; and (iii) between comparisons—1.A versus 2.B and 2.A versus 1.B. Under standard microeconomic arguments, the final comparison (iii) is generally presumed valid in the presence of ordering effects (in which 1.A (1.B) is judged to be different from 2.A (2.B)).

particular set of dependent variables. Orthogonality is critical to statistical identification of treatment effects and other parameters. Thoughtful design and careful documentation should ensure that sessions are instituted such that the only aspect that differs between the control and one or more treatments is directly related to the proposition being evaluated or some auxiliary factor that must be identified and controlled. A poor experimental design could have more than one factor changing between control and treatment, contaminating inferences that might be drawn about treatment effects.[5]

For complex experimental designs that involve multiple factors, factorial designs should be considered. A factorial design separates out each factor of interest, plus the interactions among them, into separate treatments. For example, if the treatment of interest involves exposing subjects to both A and B factors simultaneously, a factorial design would include four treatments: control, exposure to A, exposure to B, and exposure to both A and B. Factorial designs provide for thorough exploration of factors and potential interactions, but they are more costly to implement and might not be necessary if the overall effect of the factors in concert is the only treatment of interest. On the other hand, if there are no interactions among factors, only $n + 1$ treatments are necessary to identify n main effects.

Once an orthogonal design is established, it is important that subjects be randomly assigned to control and treatment conditions to preserve the comparability of subjects across cells. "Randomization ensures that the assignment to treatment is independent of other sources of variation, and that any bias is balanced across treatment and control groups, thus assuring that the estimate of the average treatment effect is unbiased" (List et al. 2011, p. 442).

The simplest randomization technique (e.g., coin flip or die roll) entails random assignment of subjects to treatment and control groups. While minimizing the risk that treatment is correlated with individual characteristics, this approach generates random sample sizes for treatments and controls and could induce a larger variance in the outcome variable than alternative procedures.

Alternatively, subjects can be systematically allocated to treatment and control groups until appropriate cell sizes have been attained. Conditioning subjects in the sample on observable factors (e.g., age, employment status) can reduce variance in the outcome variable, which will improve the power of the experiment, though this can limit the ability to generalize inferences.

In field studies, there are useful ways to introduce randomization that can be more equitable and less disruptive than standard experimental approaches (Duflo et al. 2007). Provision to eligible participants via lottery is a natural and equitable allocation mechanism if treatment offers potential beneficence (as may be the case, for example, in economic development studies) and resources are limited. In some

[5]Some experiments that use group interactions or collective decision-making can entail some loss of control in evolution of interpersonal group dynamics, which can introduce an additional dimension of heterogeneity into treatment effects (Duflo et al. 2007). If group, firm, or village interactions are an integral part of the study, clustered experimental designs can be used (Bloom 1995; Spybrook et al. 2006).

instances, financial and administrative constraints can lead to a randomized phase-in of changes in resource allocation or other programs, which can create quasi-randomization that can enable inferences to be drawn. Care needs to be taken, however, to ensure the underlying allocation process is, in fact, random and not driven by covert political or social factors.

Finally, researchers can randomize assignment of an encouragement to adopt an existing program or treatment if subscription within a population is not universal. This can work especially well if randomization is infeasible for ethical or practical reasons.

Selection of the number of subjects and allocation of those subjects to treatment and control cells influence the power of the experiment. The power of a statistical test is the probability of rejecting the null hypothesis when the alternative hypothesis is true. In the past, heuristics such as "at least 30 subjects per cell" have guided experimental design, though such designs are often less than optimal and could lack the desired power. Determination of optimal cell sample sizes requires consideration of four elements

1. The significance level (α) to be used in hypothesis tests, equal to the probability of falsely rejecting a true null hypothesis (i.e., of committing a Type I error).
2. The associated power of the hypothesis test (κ), equal to the probability of correctly rejecting a false null hypothesis (i.e., 1—probability of committing a Type II error, where a Type II error is not rejecting a false null hypothesis).
3. The minimum detectable treatment effect size (δ).
4. The standard deviation of the effects (σ; in this example, assumed known and equal across treatment and control).

The basic formula for required sample size is

$$N \geq \left[\frac{\sigma \cdot \tilde{z}}{\delta}\right]^2, \tag{10.1}$$

where $\tilde{z} = |z_\alpha + z_{1-\kappa}|$ indicates the sum of standardized z-scores associated with Type I and Type II errors,[6] and the variance is assumed to be equivalent across the treatment and control groups (Cohen 1988).

Testing for differences in the mean outcome across treatment and control can be couched in a regression framework

$$y_i = \beta_0 + \beta d_t + \varepsilon_i, \tag{10.2}$$

where y_i is the outcome variable, β_0 is the mean outcome of the control group, d_t is a dummy variable for the treatment group, and β is the average treatment effect. The residual, ε_i, is assumed to be independent and identically distributed across

[6]For example, if $\alpha = 0.05$ and $\kappa = 0.8$, then $z_\alpha = \Phi^{-1}$ $(0.05) = -1.645$ and $z_{1-\kappa} = \Phi^{-1}$ $(0.20) = -0.842$, so that $|z_\alpha + z_{1-\kappa}| \cong 2.487$.

observations, which can be violated if assignment to treatment and control groups is not random. Equation (10.2) can be expanded to include individual and other cross-sectional covariates. With a continuous outcome variable and one treatment, the variance of the estimated treatment effect is given by

$$\mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{P(1-P)N}, \tag{10.3}$$

where σ^2 is the residual variance, N is the sample size, and P is the proportion of the sample allocated to treatment (Duflo et al. 2007).[7] By convention, the typical analysis will test the null hypothesis (H_0) that the treatment effect is zero. For a given α, the null hypothesis is rejected if

$$\left|\widehat{\beta}\right| > \tau_\alpha \times \sqrt{\frac{\sigma^2}{P(1-P)N}}, \tag{10.4}$$

where τ_α is the critical value from the standard t-distribution ($\tau_{\alpha/2}$ for a one-tailed test). The power of the test for a true effect size β is the proportion of the area under the t-distribution that falls outside the critical range $-\tau_\alpha$ to τ_α; this is the probability that we reject H_0 when it is, in fact, false. To achieve power κ (e.g., 80% or 90% is commonly employed), we require

$$\left|\widehat{\beta}\right| > (\tau_{(1-\kappa)} + \tau_\alpha) \times \sqrt{\frac{\sigma^2}{P(1-P)N}}, \tag{10.5}$$

where $\tau_{(1-\kappa)}$ and τ_α can be read from a t-table. For example, if we desire a 90% power level, $\tau_{(1-\kappa)} = 1.660$ (for a sample size of 100). Thus, if $\alpha = 0.05$ (for which $\tau_\alpha = 1.984$), and we desire 90% power of rejecting the null when it is false, δ is 3.644 times the standard error of $\widehat{\beta}$.[8]

Equation (10.5) indicates that there is an inherent trade-off between the probability of committing Type I and Type II errors. Overall, the power of an experiment is determined by the desired δ, the residual variance, the proportion of the sample allocated to control and treatment, and sample size. In the simple regression framework above, variance of the treatment effect and δ are minimized by setting $P = 0.5$, allocating half of the sample to treatment and half to control. Optimal sample design varies with relative costs of data collection (subject recruitment and execution of experiment) and treatment (differential costs associated with treatment group). When costs of treatment are high relative to data collection, the ratio of

---

[7]This discussion closely follows Duflo et al. (2007).

[8]If control and treatments are applied to grouped data (e.g., villages in a developing country), power analysis must take account of intragroup correlation. See Duflo et al. (2007, pp. 31-33) for details.

treatment to control is proportional to the square root of the ratio of collection to treatment costs (Duflo et al. 2007). Thus, when treatment is relatively expensive, less of the sample is treated. For example, if treatment costs are twice that of data collection, the ratio of $P/(1 - P) \cong 0.707$, so with a sample size of 100, $P \cong 41$, with the remaining 59 serving as control.

Including covariates in Eq. (10.2) can decrease residual variance, which decreases treatment effect variance (Eq. 10.3) and the minimum detectible effect size δ, improving the power of the experiment to detect an effect if one exists. Alternatively, the researcher can stratify on observable subject characteristics, creating a blocked experimental design that implements treatment and control within each block (List et al. 2011). This will typically reduce the variance of the outcome variable and the residual variance. A weighted average of differences across blocks produces an estimate of the average treatment effect (Duflo et al. 2007).

In order to apply these methods, the researcher must form prior beliefs about the magnitude of variance before designing the experiment. Variance can differ across treatment and control groups. Knowledge of the variances can be derived from theory, pilot studies, or previous empirical results. Equations (10.3)-(10.5) present the case of a continuous outcome variable with a single treatment cell. In this case, when σ^2 is the same in both treatment and control cells, treatment and control sample sizes are equivalent. If variances are not equal across treatment and control, the fraction of the sample allocated to treatment (control) is proportional to the relative variability, $\frac{\sigma_t}{\sigma_c + \sigma_t}\left(\frac{\sigma_c}{\sigma_c + \sigma_t}\right)$. If sample sizes are large enough to justify the normal distribution as a good approximate for the t-distribution, requisite sample sizes can be calculated analytically as

$$N = \left(\tau_{(1-\kappa)} + \tau_\alpha\right)^2 \times \frac{\sigma^2}{\delta^2}, \tag{10.6}$$

if the variances are equal, and as

$$N^* = \left(\frac{\tau_{(1-\kappa)} + \tau_\alpha}{\delta}\right)^2 \times \left(\frac{\sigma_c^2}{P_c} + \frac{\sigma_t^2}{P_t}\right), \tag{10.7}$$

if the variances are unequal, where P_i is the proportion of the sample devoted to control (c) and treatment (t).[9] For small sample sizes, requisite N must be approximated numerically. List et al. (2011) advised that "If the variance of outcomes under treatment and control are fairly similar there should not be a large loss in efficiency from assigning equal sample sizes to each" (p. 446). Additional background on power analysis is found in Cohen (1988) and Montgomery (2005).

[9]Equations (10.3)-(10.5) can be adjusted for unequal variances by employing the weighted-average variance estimate in Eq. (10.7).

10.1.4 Subject Payments

Payments to subjects are used in experiments in two ways. First, monetary incentives can be employed to induce subjects to participate; this payment is often referred to as the "appearance payment" or "show-up fee." Payment for appearing at a designated place and time provides motivation for participation and can provide a signal of credibility for the researcher.

Second, and perhaps more important, outcomes associated with decision-making in the experiment are generally believed to require tangible rewards (payoffs) to induce effort and cognition on behalf of subjects. Sometimes payoffs are denominated in points or experimental dollars that convert to money or prizes at some specified rate. Such money filters can provide the researcher greater control over the payoff gradient (producing a finer or coarser scale as needed); can prevent interpersonal payoff comparisons (if payoffs are common knowledge) using unique, private conversion ratios; and could inhibit potentially problematic focal points (e.g., sufficient hourly wage rate or expected market price) in the payoff distribution (Davis and Holt 1993). Money filters, however, can also create money illusion[10] among subjects, which can inhibit saliency of payoffs or induce gamelike speculative behavior. Generally, payoffs should be denominated in actual currency and should be kept strictly confidential unless there are compelling reasons to do otherwise. In some cases commodities or services can serve as tangible rewards, but the researcher must be cognizant that such payoffs will be evaluated differently among subjects, and subjects could suffer from satiation (which is less likely with money).

It is reasonable to set the average level of payoff equal to the marginal opportunity cost of subjects' time, which will vary across subject pools. Researchers, however, must ensure that outcomes associated with different decisions have sufficiently different rewards. Payoff differentials need to compensate for cognitive effort and other individual costs associated with evaluation and decision-making in order to avoid the so-called "flat payoff problem," in which marginal rewards from optimizing decisions are insufficient to induce careful evaluation or rational choice. In economic experiments, the existence of tangible rewards that vary sufficiently across outcomes is referred to as "saliency" of incentives.

A special concern is that accumulating wealth within an experimental session can change subjects' risk-based decisions in unpredictable ways. In order to control for wealth effects, researchers sometimes randomly select one or more decisions from a series within a session for actual payment. This procedure renders a limited number of choices as salient, but since each decision presumably has equal probability ex ante, this approach is generally seen as acceptable. Of course, if used, the procedure should be made completely transparent to subjects. Risk preferences are

[10] "Money illusion" usually refers to a tendency to treat money in nominal rather than real terms, but here it refers more generally to any behavior that does not treat money as actual currency.

intrinsic to individuals and can be context-dependent. If heterogeneity of risk preferences interferes with isolating the effect of the manipulation of interest, there are procedures that can be applied to induce a class of risk preference.[11]

10.1.5 Comparing Laboratory and Field Experiments

This section provides background information on types of economic experiments relevant to nonmarket valuation and closely follows the taxonomy proposed by Harrison and List (2004). Experiments can be classified according to the nature of the subject pool, the information that subjects bring to experimental tasks, the character of the commodity or policy under consideration, the rules of engagement, the stakes involved, and the context/environment for decision-making.

Laboratory experiments employ an externally imposed set of rules, abstract framing for subjects' decisions, and a convenient subject pool. The standard subject pool for academic researchers is usually students but sometime includes university staff members (or other easily accessible persons) as well. Information, commodity/policy, stakes, and context can vary according to the nature of the lab experiment. For example, student subjects might possess significant background information on consumptive value of products from a university store but could be much less familiar with sustainable, fair-trade, or green products. This could be seen as an advantage if the researcher wishes to engage subjects with a clean slate in terms of preconceived notions of the value or utility of green products, but it is likely to be a disadvantage if the researcher is attempting to make an inference about agents with greater income and variable propensity to purchase green products. If researchers need to increase certainty about the ability to generalize their findings to a relevant population, they may need to seek a nonstandard subject pool. Recruiting outside academia could require changing the stakes to induce sufficient interest in participation among the target population.

Experimental evidence indicates that the context in which a decision is cast can have important implications for how subjects behave. This proposition violates the choice-theoretic tenet of invariance, requiring that different representations of the same choice problem yield the same implied preference structure (Tversky and Kahneman 1981); this finding is of practical import in experiment design because context can be manipulated. On the one hand, decisions can be framed as simple choices, devoid of any contextual clues that might evoke experiential cues, emotions, or heuristics.

---

[11]For example, a two-stage procedure can be employed to induce risk neutrality by (i) allowing subjects to earn points (n out of a total possible N) in round 1, and (ii) mapping the points into a binomial utility function in round 2, in which the probability of winning a large sum of money (or prize) is n/N and the probability of winning less (or no money, or a small prize) is $(1 - n)/N$. (For more details, see Davis and Holt, 1993).

On the other hand, decision-making in the laboratory can introduce real-world context. The advantage of introducing context into an otherwise sterile laboratory environment is the degree of control over which contextual elements are emphasized; the disadvantage is the potential lack of realism in how context interacts with decision-making.

Experimentation in field settings invokes elements of experimental design, albeit sometimes in different ways from in a lab experiment. In a general sense, field experiments occur in a more "natural" environment than do lab experiments. This often implies some degree of self-selection on the part of subjects; a prevalence of context, history, and individual experience; and less intrusive interactions with subjects as they make decisions than in a lab experiment.

Harrison and List (2004) distinguished among three kinds of field experiments that differ mainly in the degree of control exerted by the researcher. The first, an artefactual field experiment, is like a conventional lab experiment except that it does not employ a convenience sample of subjects but rather seeks out subjects (e.g., farmers, nonprofit managers, business CEOs) that are particularly germane for the research question at hand. Standard laboratory protocols are used, but pervasive contextual aspects that remain uncontrolled are background artifacts in this type of experiment.

Second, framed field experiments engage nonstandard subjects in a field environment, incorporating natural context in commodity, task, experience, information set, and the like, but impose direct control over the choice options available to subjects. Framed field experiments typically involve direct interaction of subjects with the experimenter or appointed field personnel. As with lab experiments, artefactual and framed field experiments can explicitly introduce additional contextual elements via framing of background information, highlighting social factors, use of positive or negative framings for choice outcomes, presenting ethical dimensions, or other aspects of experimental design.

Finally, a natural field experiment employs a natural or endogenous set of rules and incorporates naturally occurring context with nonstandard subjects. This type of experiment incorporates variation that arises naturally or occurs with only minor manipulation within the field setting to test a hypothesis about economic behavior. Subjects typically do not engage directly with experimental personnel; in natural field experiments, subjects usually do not recognize that their behavioral circumstances constitute part of an experiment. Thus, natural field experiments employ the most realistic decision-making environment.

Laboratory experiments generally offer greater control than field experiments. The typical threats to internal validity—omitted variables and simultaneity—are minimized in lab experiments through thoughtful experimental design. Mismeasurement problems can often be managed in the laboratory, and complications due to trends in outcomes across treatment groups are not a problem because the outcomes are usually experimentally derived. Selection and attrition effects can be minimized when standard recruitment procedures are used and salient incentives for participation are offered. The potential drawback of laboratory experiments is the sterility of the decision environment (and possibly the limited background of convenient subjects).

Sterility helps to ensure that proper controls are binding on the observed behavior, but it can limit external validity. Sterility implies that there is a lack of real-world complexity to the social, cultural, political, or economic context of the experimental exercise. While the script or instructions of the experiment could introduce contextual factors, it is often difficult to accurately mimic or simulate naturally occurring contexts.

Moreover, inference from interactive experiments (whether laboratory or field experiments) can be corrupted if subjects are aware that their actions are being scrutinized by others and behave differently because of it. This phenomenon, known as the "Hawthorne effect," can be lessened—although perhaps not eliminated—by appealing to subjects that there is no correct or expected answer and ensuring that experimental proctors are not known individuals who could influence subject response (e.g., an instructor or boss).

An experimental setting where the researcher has control over information can focus on inherent economic values that reside with the subject (typically referred to as "homegrown values" in the experimental literature) or can induce individual economic values. A homegrown value is essentially an individual's subjective, personal assessment of the value of a commodity or service; these types of values are the target of most every nonmarket valuation exercise and application.

Perhaps more foreign to nonmarket valuation practitioners are induced values, in which researchers assign individual subjects abstract values that they are to use in decision-making. For example, subjects in an induced-value experiment could be expected to use the value on a playing card as their consumptive value for an artificial commodity. The use of induced values obviates problems of value uncertainty and preference formation, focusing attention on the validity of the measurement protocol and on performance (e.g., efficiency) of economic institutions and incentive compatibility. Thus, induced-value experiments can play an important role in assessing the validity of nonmarket valuation procedures (under the assumption that the underlying values of interest are well-formed and readily expressed).

In contrast, homegrown values would be of primary interest when studying value uncertainty, preference discovery and formation, or the external validity of measures of economic value. From here on, the term "economic value" implicitly refers to homegrown values without explicitly stating so (as is typical in the nonmarket valuation literature).

While the lab can serve as a place for initially vetting theory, field experiments allow for extension to potentially more realistic settings. Field experiments can take advantage of the social and cultural context within which the economic decision of interest naturally occurs. The incorporation of a realistic setting and the decision-making process that naturally occurs in such a setting are of practical import for testing theory in which these aspects do in fact influence decision-making. Indeed, subjects in field experiments have self-selected into the field environment, rendering them perhaps most relevant for inference about actual socio-economic behavior. Although the natural setting and endogenous nature of the subjects impart a high degree of external validity to the experiment, that validity comes at the cost of some loss of control.

The natural variation in experience, knowledge, beliefs, perceptions, and other idiosyncratic factors that can occur in field settings can create background noise that complicates experimental testing procedures. Regarding these potential confounds, interaction with subjects within a framed field experiment allows the experimenter to collect information on experience, knowledge, beliefs, etc. Compared to a natural field experiment, a framed field experiment offers greater control in that the researcher can exercise discretion over which elements are internally or externally imposed within the experiment. Since the experiment is framed, researchers can closely monitor subjects' behavior and manage any attempts to obtain extra information, confer with other participants, or seek advice from friends. The downside is the potential for Hawthorne effects—recognizing that their behavior is being monitored, subjects might act differently than they would otherwise.

A natural field experiment, on the other hand, makes full use of naturally occurring conditions; the rules, context, commodities, and information are primarily endogenous to whatever socio-economic institutions are being observed. The chief advantage of this approach is the extant realism of the situation. The potential for Hawthorne effects is very small, but the degree of control exercised by the researcher is more limited. Choice of the proper experimental protocol requires consideration of all of these elements.

10.1.6 Pretesting and Subject Recruitment

Once an initial design is configured, the experiment should be pretested for programming fluidity (especially for computer-based experiments), subject understanding, response recording and coding, and testing of other procedural aspects. Critical design and protocol decisions should be made with a clear assessment of pros and cons while considering the possibility of unintended consequences. All procedural information and protocol decisions should be carefully documented; this will facilitate later reporting, which should include sufficient detail that a reader is able to replicate the experiment. Pretesting can lead to refinement of experiment design, protocol, or procedures; this process should iterate until a suitable experiment has been designed.

Subjects should be recruited from the population of interest, be it readily available undergraduate students, participants in a particular market, rural villagers, small-business owners, corporate CEOs, etc. Recruiting can be conducted in person, through media advertising, via word of mouth, using social networks, or any other communication method that provides broad exposure to the population of interest. Reporting the range or expected level of earnings encourages participation. Subjects should be informed upfront not only about possible earnings but also about the length of the experiment, the nature of the decision task (though this can remain vague if need be), and any risks associated with the study (as necessitated by institutional review boards, from which approval will need to be obtained by academic investigators).

Experiment sessions should be scheduled to accommodate different subjects' schedules, and subjects should typically be allocated to sessions with a randomization procedure that ensures comparability across treatment cells. Calculation of earnings and subject payments should usually be made in private unless public revelation is part of the experiment design. Any procedural irregularities should be documented.

Critics have often argued that student subjects provide an inappropriate basis for inference on economic phenomena. The veracity of this claim, however, depends on the objective of the experiment. If the objective is to test economic theory, the standard reply, "general theory should hold for everyone," applies. If we want to test a behavioral theory, why not start with a readily available subject pool? If the purpose of the experiment is to estimate a population parameter, however, subjects should be drawn from the relevant population of interest. Nonstandard subject pools (i.e., moving away from student or other easily accessible subjects) can entail a general sample of households or economic agents that have selected into a particular market or institution. The use of subjects from a particular population introduces historical context, experience, and pertinent subject knowledge into the experiment, which can increase external validity. In the laboratory or field, participant surveys can permit analysis of these pertinent factors (e.g., experience, perceived context, etc.). Natural field experiments lack these dimensions of control.

10.2 Experimental Design in Stated Preference and Revealed Preference Methods

This section provides background and examples of the role of experimental design in nonmarket valuation. The use of general experimental design features in stated preference analysis is covered first, including within and between comparisons in both contingent valuation and choice experiments. The discussion then moves to quasi-random experimental applications (differences and difference in differences) in revealed preference analysis.

10.2.1 Stated Preference

Elements of experimental design play a major role in stated preference analysis. A chief advantage of stated preference methods is the ability to introduce independent and orthogonal variation in an element that the researcher wishes to study. This could be accomplished, for example, by introducing an additional scenario in a stated preference study.[12] In designing stated preference experiments, the researcher chooses the control and treatment conditions, levels of the attributes, within or between comparisons, and an appropriate method of randomization. These considerations arise, for example, in design of contingent-valuation and contingent-behavior survey instruments (see Chap. 4). Within the context of a stated preference survey, treatments are usually described as scenarios, indicating some sort of change in conditions and how the changes are implemented and funded. Stated preference methods are very flexible; scenarios can introduce variations in environmental conditions, resource allocation, ecological services, management approaches, levels of risk, or other factors of interest, and they can provide for tests of scope and scale. Since stated preference surveys can collect information on the same sets of behaviors under various conditions, they can explicitly employ different treatments.[13]

The introduction of multiple treatments into a given survey version allows for both within and between comparisons of treatment effects. Including multiple treatments in a within-subject design, however, raises the possibility of both respondent fatigue and perceived implausibility—potential repercussions that deserve serious consideration during the design and pretesting phases. If multiple treatments are used within a particular survey version, it is of critical importance to consider changing the order of treatments so that between comparisons are preserved in the presence of order effects (see Footnote 4 for details). That is, if ordering seems to matter "within" (which is commonly found), additional use of "between" comparisons through changes in order provides for a robust research design.

In contrast to contingent valuation (Chap. 4), stated preference choice experiments (see Chap. 5) have a more direct link to experimentation because the theory of efficient choice set design is built on experimental principles. Choice experiments use the framework of discrete-choice theory (see Chap. 2) and permit respondents to choose among several distinct alternative scenarios (or "profiles") that differ by attributes (one of which is typically a cost variable). The profiles are assembled into choice sets, and a respondent typically evaluates more than one choice set (thus making multiple choices). Because the researcher designs the profiles, compiles the choice sets, and specifies the number and array of choice sets shown to a respondent, the matrix of independent variables for a conventional

[12]Such flexibility is typically absent from revealed-preference studies. For example, resource quality might not have adequate variability in a revealed-preference hedonic price or recreation demand dataset, or it might be correlated with one or more existing factors in the study. Of course, the potential downside of stated preference is that the scenarios are usually hypothetical and could lack realism or plausibility.

[13]Revealed-preference studies can also observe the same sets of behaviors under different conditions, but the conditions are typically not controlled by the researcher, or the level of control is considerably lower.

discrete-choice model (e.g., conditional logit) is entirely derived from the choice set design. As such, the design and presentation of choice sets influence the standard errors of the model parameter estimates; efficiency of the discrete-choice model is thus affected by choice-experiment design (see Chap. 5 for detailed information.)

The following two examples of experimental design in stated preference will help to illustrate some of the concepts presented above. Lew and Wallmo (2011) conducted a choice experiment using both within- and between-subject comparisons to test for scope and embedding effects in valuing protection of endangered and threatened species. Their experimental design systematically allocated subjects to one of two treatment cells: one that offered improvements in the levels of protection for two endangered or threatened marine species (Hawaiian monk seal and smalltooth sawfish), and the other, which also included protection levels for a third marine species (Puget Sound chinook salmon). The choice sets were composed of a status quo situation with no improvements in protection at no additional cost and two alternatives that provided for improved protection levels at some individual costs. Each subject was exposed to three choice sets, and thus could make three choices for the level of species protection. The combination of orthogonal treatment cells with survey designs offering different protection levels across the various choice sets (in varying order) allowed for tests of scope sensitivity (indicating whether economic value increased, or at least did not decrease, with the extent of endangered species protection) and embedding effects (where the levels of protection and number of species protected changed simultaneously), both within and between subjects.

Whitehead et al. (2008) combined revealed and stated recreation trip data in order to value beach trips and assess improvements in beach access and beach width. They collected revealed preference data on previous-year trips to southeast North Carolina beaches, as well as statements of planned number of trips over the next year if conditions were to remain unchanged. Planned future trips with unchanged beach conditions were a hypothetical baseline that provided orthogonality in combining revealed preference and stated preference data.[14] Whitehead et al. followed the measure of baseline stated preference trips with inquiries about future trips under the following enhanced conditions: (1) parking and admittance points would be increased to provide easier access and (2) beach width would increase by 100 feet. By collecting stated preference contingent-behavior

[14]The collection of planned stated preference behavior under current conditions is a method proposed by Whitehead et al. (2000). The stated reference baseline trips may suffer from hypothetical bias (optimistic assessment of future trips). Since recreation conditions are held constant in the stated preference baseline treatment, the researcher can use the stated preference data under current conditions to control for hypothetical bias or any other expected change in household conditions (expected increase in income, reduced opportunity cost of time). Measuring stated preference demand under baseline conditions before proceeding to changing conditions represents careful experimental design that avoids changing multiple factors (data type and conditions) within a single treatment (Whitehead et al.).

information from each respondent under baseline conditions and conditions described under (1) and (2), their experimental design allowed for within-subject comparisons of beach trips under current and hypothetical conditions.[15]

10.2.2 Revealed Preference

In revealed preference analysis, researchers do not have the luxury of designing their covariate matrix. Problems for inference in revealed preference analysis can occur due to endogenous variation in regressors stemming from omitted variables, selection, or simultaneity.

Omitted variables that belong in a properly specified model can lead to biased coefficient estimates for any included variables that are correlated with the omitted variable (due to correlation with the random error term). In other words, the researcher will encounter problems in inference if an important variable is not included in the regression equation because some of the influence of the omitted variable could be attributed to included variables.

Selection occurs when the behavior under study influences one or more variables that are presumed to be exogenous. For example, Parsons (1991) examined the possibility that household location choice, and thus distance from a recreation destination, could be influenced by the household's expected demand for visits to the destination. If household location choice and recreation avidity are codependent, distance from a recreation destination cannot be used as an exogenous instrument in a travel cost model to identify recreation demand (and consumer surplus).

Finally, "simultaneity" refers to endogenous variables that are jointly determined with the regression's outcome variable. In the parlance of experimental design, these endogeneity problems can be thought of as a violation of orthogonality, where some unobserved factor is correlated with cell assignment, rendering estimation of treatment effects (or other econometric parameters) biased.

In cases in which omitted variables, selection, or simultaneity present recalcitrant identification problems in revealed preference analysis, one might pursue natural experiments that take advantage of "quasi-randomization" due to some source of exogenous variation that permits estimation of key parameters. Building on concepts and findings in psychology, Meyer (1995) provided an overview of natural experiments in economics. The three main goals of a natural experiment are to (1) find a source of exogenous variation for key explanatory variables, (2) identify suitable comparison groups, and (3) estimate treatment effects and probe implications of the hypotheses under study. Exogenous variation can stem from policies or regulations that vary by political jurisdiction or by natural events or other shocks

[15] Landry and Liu (2011) provided an overview of econometric models for analyzing these types of data.

that affect some subjects, but not others. Comparison groups should provide a valid counterfactual, which can be compromised by (1) omitted variables that vary across groups and affect the outcome variable, (2) uncontrolled background trends in the outcome variable, (3) misspecified variances that ignore clustered or correlated error structures, (4) variable mismeasurement, (5) endogeneity due to simultaneity of outcomes and explanatory variables, (6) self-selection of subjects into control or treatment groups, and (7) attrition from control or treatment groups (Meyer).

The simplest natural experiment uses one group with a before-and-after (or pre- and post-) effect. (This is analogous to the basic experimental regression discussed above.) In economics, this design is often referred to as “differences” and can be described by the following equation:

$$y_{it} = \beta_0 + \beta d_t + \varepsilon_{it}, \tag{10.8}$$

where y_{it} is the outcome variable of interest for subject i in period $t = 0$ (before) or 1 (after); β_0 is the mean outcome before treatment; d_t is a dummy variable indicating pre- ($d_t = 0$) and posttreatment ($d_t = 1$); β is the true causal effect (under the assumed identifying condition) of treatment on the outcome; and ε_{it} is an independent and identically distributed error term. The identifying assumption in Eq. (10.8) is that after (or post-) observations would be equivalent to before (or pre-) observations in the absence of treatment: $E[\varepsilon_{it}|d_t] = 0$, which says that the conditional expected value of the error term does not depend on treatment assignment (also referred to as “ignorable treatment assignment”). This implies that β would be zero in the absence of treatment, which is a strong assumption in most cases because it requires that the change in the average of the outcome (dependent) variable across $t = 0$; 1 is affected only by the treatment. Thus the assumption rules out any omitted variables (contemporaneous with treatment) that affect the outcome, any background trend in the outcome, and any reverse causation in treatment and outcome. Under these conditions, $\widehat{\beta} = \bar{y}_1 - \bar{y}_0$ (the difference in means across before and after periods). The variance of $\widehat{\beta}$ is given by Eq. (10.3).

An example will help illustrate these concepts. Jeuland et al. (2010) used a natural experiment to value Mozambique households’ willingness to pay for cholera vaccine within the travel cost model framework. The cholera vaccine trials were initially designed to target only the local municipal population in the city of Beira, but once word spread that the vaccine was being offered free of charge, residents from much farther municipalities traveled—in some cases great distances—to receive treatment. The vaccine program was altered to accommodate the increased demand. Control and treatment were based on distance from Beira; the quasi-experiment was spatial rather than temporal. Households at greater distance from the location of vaccines purchased lower quantities and were less likely to undertake a trip to acquire vaccines. This application qualified as a natural experiment because an unexpected, exogenous source of variation in travel cost permitted the estimation of a travel cost demand model to value the vaccine for Mozambique households. Without this exogenous event, no travel cost model could have been estimated. The key identifying assumption was that households from other municipalities were otherwise similar to those in Beira (providing comparability across the two groups that would permit them to be included in a single model).

An approach that is likely to be more widely applicable and valid under special circumstances is the use of a before-and-after design with an untreated comparison group. Also known as "difference in differences," this approach requires a treatment group for which data are available before and after the treatment and a control group that did not receive the treatment but was exposed to some or all of the other influences that could have affected the treatment group. This design can be described by the following equation:

$$y_{it}^{j} = \beta_0 + \beta_1 d_t + \beta^1 d^j + \beta d_t^j + \varepsilon_{it}^{j}, \tag{10.9}$$

where y_{it}^{j} is the outcome variable of interest for subject i; groups are identified by $j = 0$ (control), $j = 1$ (treatment); period is given by $t = 0$ for pre-treatment, $t = 1$ for posttreatment; the dummy variables are $d_t = 0$ for pre-treatment, $d_t = 1$ for posttreatment, $d^j = 0$ for control group, $d^j = 1$ for treatment group, $d_t^j = 1$ for subjects in the treatment group, posttreatment, and 0 otherwise; and ε_{it}^{j} is an independent and identically distributed random error term. The β_0 parameter is the mean effect for the control group before treatment; β_1 is the mean time effect common to both groups (after treatment); β^1 is the time-invariant difference between the treatment and control groups; and β is the true causal effect of treatment on the outcome for the treatment group (as long as the identifying assumption is valid). The identifying assumption is that there are no omitted variables that have a posttreatment effect on the treatment group: $E\left[\varepsilon_{it}^{j} | d_t^j\right] = 0$. Again, this assumption indicates that the conditional expected value of the error term does not depend on treatment assignment. In other words, β would be zero in the absence of treatment. In this case, $\widehat{\beta} = \Delta\bar{y}_0^1 - \Delta\bar{y}_0^0 = \left(y_1^1 - y_0^1\right) - \left(y_1^0 - y_0^0\right)$, which is the difference in the outcome for the control group (pre- and posttreatment) subtracted from the difference in the outcome for the treatment group (pre- and posttreatment).

Relative to the "differences" framework, the presence of a suitable comparison control group in the difference-in-differences framework reduces the likelihood of problems due to omitted variables, mismeasurement, and background trends in the outcome variable. Comparability between treatment and control is more likely when attrition (loss of observations across the time dimension) is minimized or eliminated (Meyer 1995). The primary threat to internal validity is an omitted interaction between $t = 1$ and $j = 1$; this corresponds with an omitted variable that is simultaneous with treatment but only affects the treatment group. In both the differences and the difference-in-differences framework, including individual characteristics within a regression framework to control for the effect of other observable explanatory variables on outcomes (according to whatever underlying theory is appropriate) will generally improve the efficiency of the estimate of β and reduce the error variance.

Natural experiments employing the difference-in-differences framework have seen wide application in hedonic regression models. Hallstrom and Smith (2005) used a difference-in-differences repeat sales hedonic property model to estimate the effect of Hurricane Andrew on market values for homes in the Special Flood Hazard Area of Lee County, Florida. Since Hurricane Andrew barely missed Lee County, they contended that the damage wrought by the storm on nearby counties could serve as an information signal on flood risk in South Florida. Thus, the occurrence of Hurricane Andrew is the treatment. Homes inside the Special Flood Hazard Area are the treatment group, and homes outside the Special Flood Hazard Area are the control group. The comparability of this control group appears suitable because homes outside the Special Flood Hazard Area but in the same county are exposed to similar local labor market conditions and regional macroeconomic forces, but they face lower flood risk. Possible omitted interactions include changes in flood insurance rates (which would primarily affect homes in the Special Flood Hazard Area) after the storm; if this omitted effect were present, it could negate the assumption that $E\left[\varepsilon_{it}^{j}|d_{t}^{j}\right]=0$.

Carbone et al. (2006) expanded this framework to multiple counties in order to further examine potential threats to validity by comparing results for Lee County with those from Dade County, Florida, which suffered extensive property damage. Property damages and subsequent repairs create significant potential for omitted interactions that would complicate the use of difference in differences (negating the assumption that $E\left[\varepsilon_{it}^{j}|d_{t}^{j}\right]=0$). The authors examined a number of proxy variables to control for the omitted interactions and claimed reason for cautious optimism in their findings. Their use of multiple treatment and comparison groups provided some evidence of convergent validity of different difference-in-differences formulations in their application.

10.3 Applications of Experimentation in Nonmarket Valuation

A primary goal of nonmarket valuation is the measurement of individual or household-level, subjective economic values. In practice, this requires that subjects solve (at least) two problems: first, they must formulate a value; second, they must decide how to respond to the particular decision environment or elicitation mechanism they face (Hoehn and Randall 1987). Value formulation naturally arises within any constrained-choice problem and is part of a fundamental decision process that invokes social, psychological, and economic factors. Rational choice theory is a stylized and tractable model of this process, whereas behavioral economics and other disciplines have explored circumstances and situations that extend beyond the confines of rational choice theory.

Value elicitation involves a concerted effort to induce subjects to provide valid responses within an interview or survey protocol. Different elicitation mechanisms can extract divergent responses depending on their incentive structure and the motivation of the respondent (Hurwicz 1972). An elicitation mechanism that provides incentives for accurate (sometimes referred to as "truthful") revelation is known as "demand revealing" or, more generally, "incentive compatible." The latter term suggests correspondence between the respondents' incentives and the researchers' objectives.

Problems in nonmarket valuation can arise from uncertain values or unfamiliar institutions in which these values are to be expressed. Braga and Starmer (2005) distinguished between "value learning," in which subjects discover their own preferences, and "institutional learning," in which subjects become familiar with the rules, properties, and incentive structure of particular value elicitation protocols. This section presents several examples of the use of experimental methods to study value formation and elicitation. Perspective is provided on key decisions that the researcher had to make in designing the experiment to address the question at hand. The goal is to help the reader understand the process of experimental design, providing a foundation for developing appropriate research methods and enabling the reader to evaluate experimental results in the nonmarket valuation literature (and elsewhere).

10.3.1 Formation of Preference

Building on previous theory and experimental findings, Bateman et al. (2008) identified three fundamentally distinct conceptual frameworks describing the formation of individual preference: (1) the neoclassical microeconomic perspective, in which preferences are known by individual decision-makers, stable, and readily revealed in incentive-compatible settings; (2) the preference discovery hypothesis, in which preferences are exposed, or perhaps learned, through a process of repetition, feedback, and experience (Plott 1996); and (3) the "coherent-but-arbitrary" perspective, in which preferences are internally consistent but influenced by arbitrary signals, referred to as "anchors" (Ariely et al. 2003).[16]

Formation of preference has been studied for conventional goods and services (see, e.g., Kapteyn et al., 1980). In that context, some researchers claim that socio-psychological forces, such as peer effects or habit formation, play an important role in the evolution of preference, while some neoclassical economists tend to view preferences as egocentric and immutable. In the context of nonmarket valuation—and stated preference in particular—however, preferences can be latent or ambiguous because subjects do not routinely make constrained choices about

[16]The coherent-but-arbitrary perspective is similar to constructed preferences (Slovic 1995; Kahneman 1996), which models preferences as not prior to, but rather constructed during the decision-making process. These models have not been embraced by economists because they are antithetical to axiomatic preference modeling that is the cornerstone of microeconomic theory and welfare analysis (Braga and Starmer 2005).

nonmarket goods that would permit the expression or discovery of preference. Thus, there can be considerable uncertainty about one's willingness to pay for a change in the level of some public good or externality. Moreover, lack of repetition and experience limits the ability of subjects to learn about their preferences. Under these circumstances, people's answers to survey questions could be unreliable signals of value because they are unaccustomed to explicitly valuing public goods and externalities, and they are unfamiliar with the typical elicitation mechanisms employed in nonmarket valuation exercises. This section focuses on experimental applications in nonmarket valuation that deal with preference formation and value learning and subsequently explores issues related to institutional learning and preference elicitation.

10.3.1.1 Value Uncertainty

One of several aspects relevant to value uncertainty is the importance of reflection—having time to think about the relative importance of a commodity or service, discuss it with others, and otherwise seek out information relevant to its potential utility. Cook et al. (2007) conducted a framed, stated preference field experiment designed to test for the importance of additional response time (time to think), in the context of a survey to estimate willingness to pay for cholera and typhoid fever vaccines in Vietnam. The sample was focused on households with children under 18 and for which the head of household was less than 65 years old. By focusing their experimental design in such a way, the researchers reduced the heterogeneity of respondents, making their research design more powerful by reducing the variance of responses. Their subjects were aware of cholera and typhoid fever, but most did not know anyone who had contracted one of these diseases. Thus, the commodity was likely to be relevant for households in Vietnam, but since the vaccines were generally not available, the market price of field substitutes was not likely to unduly influence valuation responses.

In order to evaluate the influence of time to think, discuss, and seek relevant information, Cook et al. (2007) split their sample: half of the respondents answered stated preference questions during the course of a face-to-face interview, whereas the other half took the survey instrument home and completed the questions at their leisure over the course of an evening. The researchers randomized treatment assignment and conducted the face-to-face interviews first in order to avoid confounding effects due to leakage of information from time-to-think respondents onto subjects in the control group. If those in the time-to-think treatment were to communicate details about the study to subjects in the control group, inference of treatment effects could be compromised.

Randomization of respondents to the control and treatment cells allowed the researchers to make inferences about the influence of additional time, the opportunity to discuss vaccine choices with family and friends, and the ability to gather more information on the value of the goods. Note that Cook et al. (2007) did not use a factorial design to attempt to separate the influence of these factors; their experimental design is capable of only identifying an overall effect. Further, their design potentially suffers from a difference between treatment and control—the presence of the interviewer in the control sessions—which can tend to elevate willingness to pay. Results indicate that time to think reduces stated willingness to pay for vaccines, and respondents in the time to think treatment made fewer preference errors, such as reversing preference, making an intransitive choice, or violating the assumption of monotonicity (i.e., “more is better”). Other than the interviewer difference, the between-design permits an evaluation of the overall influence of time to think on household valuations of vaccines. Findings suggest time to think renders greater internal validity and lower estimates of economic value.

10.3.1.2 Value Learning

Nonmarket valuation research is typically focused on “low experience” or novel goods—that is, commodities and services that individuals are not accustomed to evaluating in terms of payment, compensation, or other forms of trade-offs. Motivated by concerns regarding empirical results of large price premiums for new products with environmental/health attributes (such as irradiated meat, pesticide-free fruit, and growth-hormone-free milk and beef), Shogren et al. (2000) implemented a laboratory experiment with consumer commodities. They hypothesized that the novelty of the experimental setting or the commodity under scrutiny could create abnormally large price premiums. Lack of familiarity with the experimental protocol could induce investigative bidding (i.e., trying out different bidding strategies because the cost of doing so is low) that creates high price premiums. On the other hand, the novelty of the particular commodities being evaluated could encourage larger bids that reflected perceived consumptive value as well as an informational premium stemming from learning how the good might fit into the subject’s preference set.

To obviate any hypothetical aspects of the experiment, subjects were recruited to participate in a multisession experiment that would offer monetary rewards and free lunches. Thus, Shogren et al. (2000) endowed the subjects with commodities initially and evaluated willingness to pay to “upgrade” to an alternative commodity. Subjects must agree to consume the lunch within a session before being paid their appearance payment. To explore the implications of increasing familiarity with decision-making in experiments and consumption of common or uncommon food products, they conducted sessions over two or three weeks. In each session, subjects bid in sealed-bid, second-price (Vickrey) auctions, which were theoretically incentive-compatible because they decoupled the probability of winning from the price paid. Their tasks involved trading an endowed good (name-brand candy bar, conventional pork sandwich, and apple) for an alternative good (different candy bar, irradiated pork sandwich, and mango, respectively); the rationale for endowing subjects with a particular good and auctioning an exchange for an alternative good is that it creates a common consumption baseline across subjects and limits overzealous competition for the auctioned item. Information was provided on health risks associated with non-irradiated meat so that subjects were fully informed about latent risks of pork consumption and the benefits of irradiation.

The experiment utilized repeated bidding sessions over multiple lunches in order to assess the influence of gaining familiarity with the experimental setting. If novelty of setting leads to exploratory bidding that inflates prices, bids should decrease over time as subjects become familiar with the experiment. If, however, lack of familiarity with food products is the source of abnormally high price premiums, one might expect to see decreasing bids for only unfamiliar food products, particularly for those who have gained experience with those goods. Of note, food commodities in the experiment were selected to vary systematically in terms of subject familiarity—from most familiar to less familiar to least familiar; this sequencing was not varied across subjects or over time. A shortcoming of this experimental design is that if there were order effects, they would complicate interpretation of the results. Shogren et al. (2000) conducted five rounds of bidding for each commodity, with one auction for each commodity randomly chosen as binding (and thus being executed during the experiment). Repeated auctions were chosen to permit subjects to gain experience with the auction mechanism. The identification numbers for the highest bidder and the second-highest price were posted after each round.[17] With a recurrence of five bidding sessions for three goods, four times over the course of two or three weeks, each subject provided a total of 60 observations over the entire experiment.[18]

Using a lab experiment, Shogren et al. (2000) exercised a high degree of control over context, information, and sequencing of decisions. In implementing multiple sessions over two- or three-week periods, they permitted a test of evolving familiarity with the auction mechanism, but with a loss of some control because subjects could gain familiarity with the commodities outside of the experiment.

The Shogren et al. (2000) experiment is a good example of an effective within design; the hypotheses under consideration involve novelty of valuation context and commodity, so subjects needed to repeat tasks in order provide evidence on operative dynamics. The primary comparisons are between products, but within subjects over time. Over the course of four sessions, Shogren et al. found a decaying price for the most novel of food items (irradiated meat), while mean bids for common food items were consistent across the sessions. Thus, their results suggested that the product type could be more important in creating a price premium than the auction mechanism (when learning via repetition was permitted).

[17]Corrigan et al. (2012) noted some of the controversies surrounding repeated auctions with price feedback. While this design allows for learning about preferences and auction institutions, it can engender value interdependence, anchoring on posted prices, detachment from the auction, and competition among participants.

[18]To minimize attrition, Shogren et al. (2000) provided an additional incentive payment to subjects who participated in all sessions.

Moreover, those subjects who won the auctions bid significantly lower in later auctions, suggesting that the novelty effect wears off over time as subjects become familiar with the goods.[19]

10.3.2 Elicitation of Preferences

Equally important for nonmarket valuation but formally distinct from preference formation, discovery, and learning are the effects, intended or otherwise, of methods used to elicit preferences. Carson and Groves (2007) argued that many researchers take for granted that respondents grasp the valuation questions to which they are exposed. Their nuanced examination of valuation findings revealed many compelling reasons to expect differences in patterns of responses stemming from the incentive structure or interpretation of survey questions by respondents. Elements of the stated preference design—in particular the implementation frame (specifying prospective rules for public good provision) and elicitation frame (the particular format of the valuation questions)—can influence estimates of economic value (Carson et al. 2001; McFadden 2009).

A clever paper by Bateman et al. (2008) devised an experimental protocol that enlightens the underpinnings of preference formation and elicitation. The "learning design contingent valuation" approach compared single-bounded and double-bounded, dichotomous-choice, contingent-valuation estimates of willingness to pay for improvements in public goods, both within a sample that conducted repeated valuations and between samples that conducted one versus multiple valuations. They recruited 400 households with registered voters in Northern Ireland and conducted face-to-face interviews to assess stated willingness to pay for improvements in living conditions for common farm animals. Two hundred of the randomly chosen households completed a series of double-bounded valuation tasks, whereas the other 200 households completed only one double-bounded valuation task.

Bateman et al. (2008) were interested in examining two problematic phenomena associated with value elicitation. First, anchoring occurs when subjects employ the anchoring-and-adjustment heuristic in responding to valuation or judgment questions (Tversky and Kahneman 1974). If subjects hold uncertain or ambiguous value beliefs, they can anchor their assessment on provided information (even when it is normatively irrelevant) creating a correlation between the anchor and subject

[19]Consistent with the idea of value learning, Holmes and Boyle (2005) and Kingsley and Brown (2010) found that the variability of the error component of the random utility model decreased with the number of choices a subject made; subjects appeared to better discriminate among multiattribute choice sets as they gained experience with evaluating and selecting their preferred option (DeShazo and Fermo 2002). Among an extensive series of pairwise choices, Kingsley and Brown also found that the probability of an inconsistent choice decreased with choice experience.

responses.[20] And second, individual responses to single-bounded and double-bounded valuation questions are often found to be inconsistent, suggesting a problem with preference stability or the valuation protocol. The single-bounded format is relatively inefficient because it requires a large number of observations to identify the distribution of economic values. While double-bounded (Hanemann et al. 1991) and multiple-bounded variants (Langford et al. 1996) increase efficiency, the implied value distributions from the various questions are often found to be internally inconsistent (see, e.g., Bateman et al., 2001).

Bateman et al. (2008) devised an experimental design that allowed them to draw inferences about preference formation and institutional learning. Through judicious design, they chose four farm animals that are formally distinct, yet conceptually similar: laying hens, poultry chickens, dairy cows, and pigs. They reasoned that laying hens and poultry chicken are more similar to each other than either is to dairy cows or pigs, while cows and pigs are also somewhat similar, yet distinct from laying hens and chickens. Their public good was improvements in living conditions for these farm animals.

In their within-subjects comparisons (those based on the 200 respondents who valued all four policies), they explored the influence of repetition on anchoring effects, hypothesizing that repeated exposure to a valuation protocol might induce value and institutional learning (Braga and Starmer 2005). Building on the evidence of inconsistency in single-bounded and double-bounded responses, they analyzed the influence of the initial price (randomly assigned) used in the first valuation question on the response to the follow-up question. (This is a standard approach to demonstrate anchoring in dichotomous-choice valuation responses.) They hypothesized that value learning would occur over the course of valuing the four animal types, with a possible restart in learning as the subjects move from birds (questions 1 and 2) to large, hoofed farm animals (questions 3 and 4).

Their results indicated a correlation between willingness to pay and the initial bid value for the valuation of laying hens (question 1) and dairy cows (question 3), but not poultry chickens (question 2) or pigs (question 4). This pattern of results was consistent with value learning, in which there could be some uncertainty of the value for improvement for the initial animal, but less so for the second, similar animal. Moreover, the reappearance of the anchoring effect in the valuation of dairy cows suggested that value learning might have restarted in the series of contingent-valuation responses, moving from birds to large, hoofed animals, but re-emergent uncertainty disappeared in subsequent valuation of large, hoofed animals (question 4).

To complement this result, they also showed that a between comparison with results for subjects that completed only one valuation—pigs (same as question 4 in the previous subsample)—exhibited evidence of anchoring. A random assignment

[20] A number of researchers have produced experimental findings in stated preference analysis that are indicative of anchoring effects (e.g., Boyle et al. 1985; Holmes and Kramer 1995; Green et al. 1998).

of households to the multiple- or single-valuation treatments implied that their latent preferences should be equivalent across the two treatments at the initiation of the experiment. Thus, the lack of evidence of anchoring for the experienced subjects (completing three double-bounded valuations before evaluating the improvement in living conditions for pigs) has implications for value (and institutional) learning. A series of double-bounded valuations could allow subjects to discover their preferences (as postulated by Plott 1996), whereas those who evaluated the welfare of pigs as their first valuation exercise had not had such opportunity, so they fell victim to anchoring.

To test for institutional learning with regard to the double-bounded protocol, Bateman et al. (2008) compared mean willingness-to-pay estimates from the single-bounded and double-bounded responses. Their null hypothesis was that the differences are not different from zero—response patterns were consistent across single-bounded and double-bounded referenda (in contrast with common findings). Their pattern of results is striking: for the initial single-bounded/double-bounded comparison (valuation of hen welfare for those subjects making multiple valuations and valuation of pig welfare for those making a single valuation), they found large and statistically significant differences in the welfare estimates derived from the single-bounded and double-bounded protocols, corresponding with a decrease in willingness to pay of approximately 20-42% when derived from the double-bounded model. When looking at differences in willingness to pay across the other valuations (for those subjects evaluating multiple scenarios), however, the differences across single-bounded and double-bounded models were small (ranging from 7% to less than 1%) and not statistically significant. This suggests that significant upward bias and uncertainty were present in the initial valuation. These results clearly indicate that inconsistencies between single-bounded and double-bounded responses can decrease dramatically as respondents gain experience with the double-bounded protocol. The results are robust in that they provide consistent evidence of institutional and value learning both within respondents and between samples of different respondents. It is not clear, however, that the experimental design can isolate value learning from institutional learning; each phenomenon is dynamic, arising from repetition. Moreover, the paper lacked evidence of external validity because no actual payments were made for comparison.

The next section provides further experimental analysis of preference elicitation in nonmarket valuation.

10.3.2.1 Hypothetical Bias

Survey instruments can suffer from poor design, subject confusion, or recall bias. But even when situations are adequately and realistically described, survey data can suffer from hypothetical bias. This term has witnessed broad application in the valuation literature, and many definitions have been used. List and Gallet (2001) “refer to hypothetical bias as the difference between hypothetical and actual statements of value, where actual statements of value are obtained from experiments with real economic commitments" (p. 243). List and Gallet, Murphy et al. (2005), and Loomis (2011) offered meta-analyses and summaries of results on hypothetical bias.

In some cases, differences between hypothetical and actual statements of value are believed to be related to insufficient motivation of subjects to invest time and effort to fully comprehend questions, scenarios, prospective goods, and proffered changes in goods, services, or externalities (Loomis and Ekstrand 1998; Alberini et al. 2003). Other possible explanations of hypothetical bias relate to lack of experiential context (reflecting insufficient value learning), confusion over methods and protocol (reflecting insufficient institutional learning), or subject disengagement.[21] Economic experiments have been particularly useful in exploring value formation (discussed in the previous section) and elicitation issues; a properly designed experiment can employ both hypothetical (stated preference) and real (revealed preference) valuation treatments. This is a key area where experiments have been used, providing significant insight into external validity of nonmarket valuation elicitation methods.

Within the confines of a valuation experiment, one can employ commodities with purely private consumption values (e.g., coffee mugs, paintings, chocolates) or nonexclusive public value (e.g., contributions to charity, tree planting), with the commodities potentially deliverable within the experimental setting. This is important: The potential to deliver commodities within an experiment allows for contrast of real and hypothetical commitments. In this context, one can distinguish between stated preferences, in which responses do not entail an actual economic commitment on behalf of the subject or real consequences in terms of the delivery of a good or service, and revealed preferences, in which responses entail actual commitment and/or consequences.

In using common private goods to test theories of value elicitation, researchers can encounter a number of problems related to the availability of goods in the field. First of all, value statements in this context could be censored due to the existence of substitutes in the field (i.e., even if one values a good more than the market price, why pay substantially more in an experiment when earnings can be used to purchase the good at the market price?). Second, value statements can suffer from "value affiliation" around perceived market prices (a phenomena in which bids are correlated across time as subjects make inferences from other subjects' bids, with particular implications for value formation). Finally, such statements can reflect attempts at arbitrage across experiment and field (i.e., buy low in the experiment, sell high outside of the experiment). Thus, one must take great care in making inferences from differences in stated and real economic commitments. The use of

[21]Hypothetical bias can be distinguished from strategic bias, which stems from perceived or actual perverse incentives in stated preference protocol (a fundamental reversal of incentive compatibility). Such perverse incentives for subject response can result from poor protocol design (in which subjects' optimal response is an untruthful statement), subject misperceptions that lead them to respond strategically to an otherwise well-designed protocol, or other artifacts of stated preference design that could lead to inadvertent strategic response.

contrived public goods (i.e., public goods created within the context of an experiment), on the other hand, can also have problems stemming from scale and context dependencies, the possibility of public goods being provided by others, and free riding on provision of public goods. Despite these complications, controlled experiments provide a natural testing ground for hypothetical bias in nonmarket valuation.

To assess hypothetical bias and procedures designed to attenuate this bias in contingent-valuation voting referenda, Landry and List (2007) conducted a framed field experiment at a sports card market. Their study constituted a framed field experiment because it engaged subjects who have self-selected in the sports card market as a field setting, but it exercises control over the commodity, task, and provision of goods. The primary component of interest for our purposes is the between design that employs four types of public good voting referenda.[22] Subjects were intercepted as they entered the sports card show and informed of the potential to participate in an experiment in which they would be paid $10 for showing up. The experimental sessions were held in a private meeting room located within the conference center and eight treatments were scheduled over the course of two days. Upon agreeing to participate, subjects were randomly allocated to one of the eight treatments (with the only constraint being whether they could show up at the randomly selected time). Sample sizes for each treatment ranged between 29 and 37 subjects participating in a one of the four referendum types: hypothetical, "cheap talk," "consequential," and real.

Since Landry and List (2007) were testing for the presence of hypothetical bias, the real treatment, in which subjects' responses to referendum questions had direct bearing on economic commitment and provision of goods, can be viewed as the control, while the hypothetical, "cheap talk," and "consequential" referenda were the treatments. In the hypothetical treatment, subjects' responses to referendum questions had no bearing on economic commitments or provision of goods, and subjects were made well aware of this before voting occurred. The "cheap talk" design (Cummings et al. 1995; Cummings and Taylor 1999) involves the use of a script, administered before the valuation question, which draws respondents' attention to the phenomenon of hypothetical bias and encourages them to answer the valuation questions as if they were real.[23] The moniker "cheap talk" reflects the fact that the script is nonbinding—it only encourages subjects to provide more accurate or meaningful responses.

[22]In addition, two provision prices—high and low—were evaluated, with sequencing of the prices varied randomly among treatments. The rationale for two prices was threefold: to provide more information on preferences, to increase the domain for inference of willingness to pay, and to test for preferences consistent with the theory of demand. For treatments that involved actual provision, a coin flip determined which price level would be executed.

[23]The language of the cheap talk script followed the original text in Cummings and Taylor (1999) with necessary changes for the nature of the good and provision mechanism.

The "consequential" design (Cummings and Taylor 1998), on the other hand, seeks to highlight the potential consequences of the survey's findings. If the survey has the potential to influence policy and decision-making, then respondents have an incentive to take it seriously, whereas economic theory makes no prediction about how people might respond to clearly inconsequential surveys (Carson and Groves 2007).[24] The "consequential" design is often employed in field applications of contingent valuation, where subjects are informed that results of the study will be shared with policymakers and could influence public good provision decisions. To simulate potential consequences, the result of the consequential referendum was taken as advisory, where a coin flip would determine whether or not the advisory would be binding on the subjects. Thus, referendum votes were consequential with 50 percent probability.

In their framed field experiment, Landry and List (2007) made use of private consumption goods that fit the context of their field setting. Using ticket stubs from a record-setting National Football League sporting event,[25] the researchers employed a context-specific good that subjects would expect to be traded in the field setting and for which subjects could have well-established values. Moreover, by introducing a good that was not readily available in the marketplace, they obviate censoring of responses due to the existence of field substitutes. Following Carson et al. (2001), they used private goods to simulate a public good; their binary referendum mechanism operated on a simple majority vote regarding purchase of *n* identical ticket stubs, where *n* corresponded with the number of subjects in the session.[26] Votes were recorded in private and tallied by the experiment proctor in full view of the subjects. If a majority of subjects voted to fund the public good, each received a ticket stub and each had to pay the offer price (out of their $10 endowment) corresponding with the vote. If the majority decided not to fund the public good, no one received a ticket stub and no one had to pay. This procedure was described to subjects in all treatments but was only implemented in the real treatment and the consequential treatment if the subsequent coin flip determined that the consequential referendum was binding.

---

[24]Potential consequences stemming from survey responses have been shown to produce referendum voting patterns that accord with actual voting patterns in both the lab (Cummings and Taylor 1998; Vossler and Evans 2009) and field (List et al. 2004).

[25]The ticket stubs, dated October 12, 1997, corresponded with the game in which Barry Sanders passed Jim Brown for the No. 2 spot in NFL all-time rushing yardage. One of the study's authors collected the ticket stubs after the game.

[26]While this is a clever way to simulate a public good, there could be issues creating divergence between these quasi-public goods and more typically public goods in the field. In particular, all potential beneficiaries were clearly present in the framed field experiment and the scope of the public good (in terms of both payment and provision) is clearly defined. The authors are unaware of any research that has explored the relevance of these dimensions in the context of mitigation hypothetical bias.

Landry and List found that hypothetical and real voting behaviors are substantially and statistically different, exhibiting differences of 50% points. Their results suggested, however, that both cheap talk and consequential designs can be effective at attenuating this hypothetical bias (though results in support of the consequential design are more compelling). Of note, however, the consequential design used in this experiment employed a known and relatively high probability of influence. In the field, the perceived probability of influence could be highly variable and, in many cases, relatively low. Nonetheless, this paper demonstrates the usefulness of simulated public good provision in providing for comparison of real, hypothetical, and other kinds of economic commitments.

10.3.2.2 Incentive Compatibility

Vossler et al. (2012) built and tested theory on incentive compatibility in stated preference choice experiments. They extend voting theory for single-shot, binary referenda (Gibbard 1973; Satterthwaite 1975; Carson et al. 1997; Carson and Groves 2007) to repeated binary choices that pit the status quo versus an alternative policy. Vossler et al. showed that incentive compatibility in this context can be attained if at most one policy can be implemented, the policy function (which translates votes into probabilistic action) maintains independence between choice sets, and there is a one-to-one correspondence between projects described in the survey instrument and potential policies (that is, there are no broader policy implications stemming from project voting). They expressed doubts that field applications of a series of binary discrete choices can maintain independence across choices sets and a one-to-one correspondence. Nonetheless, they implemented a field experiment to test their propositions.

Vossler et al. (2012) implemented four treatments of choice experiments that incorporated the format of their theory—pitting a proposed public good project (tree planting) against the status quo (no tree planting). The tree planting projects varied by dimension (length and width of the planting area), location (riparian or roadside), and cost, and each subject evaluated 12 binary questions (proposed project vs. no project, with ordering varied across subjects). Their experimental design consisted of three actual payment treatments involving real payments and the actual provision of tree plantings, and one stated preference treatment that did not involve real payments or actual provision.

Since independence across choice sets was a critical element of their theory, an important dimension of the experimental design varied the lucidity of independence. The actual payment treatments employed by Vossler et al. (2012) varied by clarity of independence across choice sets in provision of tree planting.[27] The

[27]Field surveys that involve choice experiments often implore respondents to treat choice sets as if they are independent. In practice, however, it is often unclear whether these exhortations are effective.

baseline treatment employed an independent lottery provision process in which the roll of a 12-sided die determined which choice set was implemented. The other treatments employed less clarity in independence across choice sets, where the provision rules were designed to encompass an array of implicit and explicit approaches used in choice-experiment field studies. Thus, the actual payment treatments varied by opaqueness of sufficient theoretical conditions for incentive compatibility.

The stated preference treatment had no stated provision rule, but all treatments included a statement indicating that results would be provided to a government agency. Thus, the experimental design permits two types of "between-subject" comparisons: real decisions for local public goods under different levels of perception of independence across choice sets and real and stated decisions for local public goods when no provision rule is explicitly stated.

Since actual and stated preference treatments vary according to actual payment and provision, Vossler et al. (2012) varied their show-up payment to compensate for potential out-of-pocket expenses in the revealed preference treatments. While seemingly minor, this could be an important design consideration to ensure that income effects across the treatments are comparable. Subjects were recruited from a mailing list of employees and friends of the academic institution that hosted the experiment and were randomly assigned to one of the four treatments. The researchers used photographs of tree planting projects to demonstrate the types of projects that could be implemented as a result of the experiment. In order to explore potential problems with incentive compatibility, they inquired about subjects' strategic motivations and perceived consequences of survey responses when shared with the government agency (in particular, relating to the likelihood that responses could affect public decisions regarding tree planting).

Vossler et al. (2012) found that willingness-to-pay functions for their four treatments varied in the following ways: (1) standard deviation of willingness to pay was lowest for the actual payment voting mechanism with independent provision, about 50% greater for the actual payment and stated preference votes that did not specify a provision rule, and about 70% greater for the actual payment votes with a provision rule in which independence was less clear; and (2) willingness-to-pay functions for the actual payment treatments were statistically indistinguishable from one another but different from the stated preference (hypothetical) willingness-to-pay functions (all based on between-subject comparisons). When they focused their analysis on respondents who viewed their choices as consequential (as measured by post-experiment queries), however, the willingness-to-pay functions were statistically indistinguishable.

In field applications of choice experiments, Herriges et al. (2010) and Vossler and Watson (2013) arrived at similar findings. Most more-recent applications of stated preference usually collect information on perceived consequences of the survey (after response to the valuation questions). Experimental evidence suggests that these perception scales can be used to calibrate responses to stated preference choice experiments; this remains an important area for future research.

10.3.2.3 Order Effects

Order effects in elicitation of nonmarket values were first explored with regard to double-bounded, dichotomous-choice, contingent-valuation responses. DeShazo (2002) postulated that subjects' initial response in a double-bounded question could shift their reference point; those who respond affirmatively to the initial price might feel as though they have completed an informal transaction, which can change their endowment. Those who decline the initial price experience no such change. This endogenous framing effect could explain the inconsistencies between value estimates from single- and double-bounded questions. Uncertainty over reservation value and the resulting effects of anchoring are another explanation for inconsistent responses across price sequences in use of double-bounded questions (Boyle et al. 1985; Holmes and Kramer 1995; Green et al. 1998). Finally, multiple-valuation questions can create impressions of a bargaining situation or convey the perception of a change in quality or quantity (Carson et al. 1997; Carson and Groves 2007), which could invoke strategic responses (Day and Prades 2010).

Field applications and experiments have also found evidence of ordering effects in choice experiments. Holmes and Boyle (2005) found spillover effects in lead and lag choice tasks, reflecting relative differences in price and commodity attributes across choices. In their application, estimates of compensating variation were sensitive to changes in the marginal utility of money stemming from modeling of sequencing effects.

Day and Prades (2010) designed a stated preference choice experiment to test for sequencing effects in individuals' responses to changes in price and commodity attributes across choice sets. Their experiment is an excellent example of careful design of treatments that allows for between comparisons of choice tasks where the only difference is the order in which particular tasks were presented to a subject. In some cases, the quality of attributes decreases across the sequences of choices, while in others it increases. In other cases, the offered price of the good increases across the sequence, while in others it decreases. The cleverly designed ordering treatments within the context of orthogonal cells permit simple and transparent nonparametric tests of neoclassical propositions, as well as ordering anomalies.

Day and Prades' (2010) application was to valuation of individual health outcomes in Spain. They designed a choice experiment that evaluated treatment of an ill-health episode, where the status quo is a free hospital treatment that does not completely eliminate symptoms (but ultimately cures the patient); the alternative is a pharmacy remedy that is more expensive, but also more effective. Recruiting 500 subjects from the general population, Day and Prades allocated them to one of six treatments, and conducted personal interviews. The essence of their experimental design was to expose subjects in separate cells to some of the same valuation questions, but in differing sequences within a series of three choices.[28]

The six treatments permitted between tests of sequencing effects (worsening and improving sequences for price and commodity) on subsequent choice questions. Day and Prades (2010) found that the frequency of choosing the alternative treatment was negatively affected by increasing price sequences, but diminishing price sequences appeared not to influence choice frequencies. They also found sequencing effects in commodity changes across choice sets, in which worsening of attribute levels resulted in reduced frequency of choice of the alternative remedy, and improvements in attribute levels produced an increased frequency of alternative choice. The results were consistent with the price framing found by DeShazo (2002) in dichotomous-choice contingent valuation and with shifting reservation prices, as suggested by the anchoring-and-adjustment heuristic. These results cast aspersion on the assumption (typically made by researchers) that responses are independent across choice sets; implications remain to be explored.[29] This fundamental insight was gleaned by an experimental design that permitted comparisons of identical choice sets appearing in different sequences across independent treatments.

10.4 Conclusions

This chapter has explored the role that experimentation can play in valuation of nonmarketed goods and services. The chapter was designed to provide the reader with an understanding of the elements of experimental design and to highlight some of the important considerations in applying experimentation in valuation. Central to the design of experiments are the issues of control and replication. The level of control can vary by the type of experiment—be it the lab, field, or some variant—but the researcher always seeks adequate control for identifying treatment effects or other relevant phenomena. Proper application requires an understanding of orthogonality, randomization, sample and cell sizes, and between or within methods of inferences. Essential to scientific inquiry using experimentation is clear documentation of all methods and procedures, which permits evaluation of results, replication of findings, and further refinement of procedures to explore other issues related to the study.

---

[28]The six treatments allowed for between tests of price and quantity effects (more commodity at the same price, and same commodity at different prices), as well as within tests of consistency in individual choices; these tests are based on neoclassical theory.

[29]Day and Prades (2010) employed stated preference methods. Thus, they were unable to compare their results with behavior in which actual payment or provision occurs. Moreover, their commodity is a private good, unlike the commodity valued by Vossler et al. (2012). Obviously, the role of independence across choice sets, provision rules, and nature of the commodity deserves further exploration.

While elements of experimental design have played a role in welfare economics in the past, many researchers that apply valuation methods have not been trained in experimentation. This chapter represents an attempt to introduce the reader to basic concepts and to draw attention to their application. Decisions regarding the setting of a study (lab vs. field), subject pool, types of goods evaluated, prevalence of context, types of values measured ("homegrown" vs. induced), method of inference (within vs. between), and elicitation/response format can have important ramifications for results and study conclusions. Researchers and practitioners of non-market valuation can greatly benefit from a better understanding of experimental design.

References

Alberini, A., Boyle, K. & Welsh, M. (2003). Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management, 45, 40-62.

Ariely, D., Loewenstein, G. & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. Quarterly Journal of Economics, 118, 73-105.

Bateman, I. J., Langford, I. H., Jones, A. P. & Kerr, G. N. (2001). Bound and path effects in multiple-bound dichotomous choice contingent valuation. Resource and Energy Economics, 23, 191-213.

Bateman, I. J., Burgess, D., Hutchinson, W. G. & Matthews, D. I. (2008). Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. Journal of Environmental Economics and Management, 55, 127-141.

Berry, J., Fischer, G. & Guiteras, R. P. (2015). Eliciting and utilizing willingness to pay: Evidence from field trials in northern Ghana. CEPR Discussion Paper No. DP10703. Retrieved from: http://ssrn.com/abstract=2630151.

Bin, O. & Landry, C. E. (2013). Changes in implicit flood risk premiums: Empirical evidence from the housing market. Journal of Environmental Economics and Management, 65, 361-376.

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. Evaluation Review, 19, 547-556.

Boyle, K., Bishop, R. & Welsh, M. (1985). Starting-point bias in contingent valuation bidding games. Land Economics, 61, 188-194.

Braga, J. & Starmer, C. (2005). Preference anomalies, preference elicitation and the discovered preference hypothesis. Environmental and Resource Economics, 32, 55-89.

Camerer, C. (1995). Individual decision making. In J. H. Kagel & A. E. Roth (Eds.), The handbook of experimental economics (pp. 587-703). Princeton, NJ: Princeton University Press.

Carbone, J. C., Hallstrom, D. G. & Smith, V. K. (2006). Can natural experiments measure behavioral responses to environmental risks? Environmental and Resource Economics, 33, 273-297.

Carson, R. T., Flores, N. E. & Meade, N. F. (2001). Contingent valuation: Controversies and evidence. Environmental and Resource Economics, 19, 173-210.

Carson, R. T. & Groves, T. (2007). Incentive and informational properties of preference questions. Environmental and Resource Economics, 37, 181-210.

Carson, R. T., Groves, T. & Machina, M. J. (1997). Stated preference questions: Context and optimal response. Paper presented at the National Science Foundation Preference Elicitation Symposium, University of California, Berkeley.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, J., Whittington, D., Canh, D. G., Johnson, F. R. & Nyamete, A. (2007). Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. Economic Inquiry, 45, 100-114.

Corrigan, J. R., Drichoutis, A. C., Lusk, J. L., Nayga, R. M. Jr. & Rousu, M. C. (2012). Repeated rounds with price feedback in experimental auction valuation: An adversarial collaboration. American Journal of Agricultural Economics, 94, 97-115.

Cummings, R. G., Harrison, G. W. & Osborne, L. L. (1995). Can the bias of contingent valuation surveys be reduced? Evidence from the laboratory. Working Paper, Policy Research Center, Georgia State University, Atlanta.

Cummings, R. G. & Taylor, L. O. (1998.) Does realism matter in contingent valuation surveys? Land Economics, 74, 203-215.

Cummings, R. G. & Taylor, L. O. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. American Economic Review, 89, 649-665.

Davis, D. D. & Holt, C. A. (1993). Experimental economics. Princeton, NJ: Princeton University Press.

Day, B. & Prades, J.-L. P. (2010). Ordering anomalies in choice experiments. Journal of Environmental Economics and Management, 59, 271-285.

DeShazo, J. R. (2002). Designing transactions without framing effects in iterative question formats. Journal of Environmental Economics and Management, 43, 360-385.

DeShazo, J. R. & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. Journal of Environmental Economics and Management, 44, 123-143.

Duflo, E., Glennerster, R. & Kremer, M. (2007). Using randomization in development economics research: A toolkit. Discussion Paper No. 6059. Centre for Economic Policy Research, London, UK.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. Econometrica, 41, 587-601.

Green, D., Jacowitz, K. E., Kahneman, D. & McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. Resource and Energy Economics, 20, 85-116.

Hallstrom, D. & Smith, V. K. (2005). Market responses to hurricanes. Journal of Environmental Economics and Management, 50, 541-561.

Hanemann, M., Loomis, J. & Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. American Journal of Agricultural Economics, 73, 1255-1263.

Harrison, G. W. & List, J. A. (2004). Field experiments. Journal of Economic Literature, 42, 1009-1055.

Herriges, J., Kling, C., Liu, C.-C. & Tobia, J. (2010). What are the consequences of consequentiality? Journal of Environmental Economics and Management, 59, 67-81.

Hoehn, J. P. & Randall, A. (1987). A satisfactory benefit cost indicator from contingent valuation. Journal of Environmental Economics and Management, 14, 226-247.

Hoffmann, V. (2009). Intrahousehold allocation of free and purchased mosquito nets. American Economic Review, 99 (2), 236-241.

Hoffmann, V., Barrett, C. B. & Just, D. R. (2009). Do free goods stick to poor households? Experimental evidence on insecticide treated bednets. World Development, 37, 607-617.

Holmes, T. P. & Boyle, K. J. (2005). Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Economics, 81, 114-126.

Holmes, T. P. & Kramer, R. A. (1995). An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation. Journal of Environmental Economics and Management, 29, 121-132.

Hurwicz, L. (1972). On informationally decentralized systems. In C. McGuire & R. Radner (Eds.), Decision and organization: A volume in honor of Jacob Marschak (pp. 297-336). Amsterdam, Netherlands: North Holland.

Jeuland, M., Lucas, M., Clemens, J. & Whittington, D. (2010). Estimating the private benefits of vaccination against cholera in Beira, Mozambique: A travel cost approach. Journal of Development Economics, 91, 310-322.

Kagel, J. H. (1995). Auctions: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), The handbook of experimental economics (pp. 501-585). Princeton, NJ: Princeton University Press.

Kagel, J. H. & Roth, A. E. (1995). The handbook of experimental economics. Princeton, NJ: Princeton University Press.

Kahneman, D. (1996). Comment on Plott's rational individual behavior in markets and social choice processes: The discovered preference hypothesis. In K. Arrow, E. Colombatto, M. Perleman & C. Schmidt (Eds.), Rational foundations of economic behavior (pp. 251-254). London: Macmillan.

Kapteyn, A., Wansbeek, T. & Buyze, J. (1980). The dynamics of preference formation. Journal of Economic Behavior & Organization, 1, 123-157.

Kingsley, D. C. & Brown, T. C. (2010). Preference uncertainty, preference learning, and paired comparison experiments. Land Economics, 86, 530-544.

Landry, C. E. & List, J. A. (2007). Using ex ante approaches to obtain credible signals of value in contingent markets: Evidence from the field. American Journal of Agricultural Economics, 89, 420-432.

Landry, C. E. & Liu, H. (2011). Econometric models for joint estimation of RP-SP site frequency recreation demand models. In J. Whitehead, T. Haab & J.-C. Huang (Eds.), Preference data for environmental valuation: Combining revealed and stated approaches (pp. 87-100). New York, NY: Routledge.

Langford, I. H., Bateman, I. J. & Langford, H. D. (1996). A multilevel modelling approach to triple-bounded dichotomous choice contingent valuation. Environmental and Resource Economics, 7, 197-211.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), The Handbook of Experimental Economics (pp. 111-193). Princeton, NJ: Princeton University Press.

Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives, 21 (2), 153-174.

Lew, D. K. & Wallmo, K. (2011). External tests of scope and embedding in stated preference choice experiments: An application to endangered species valuation. Environmental and Resource Economics, 48, 1-23.

List, J. A., Berrens, R., Bohara, A. & Kerkvilet, J. (2004). Examining the role of social isolation on stated preferences. American Economic Review, 94, 741-752.

List, J. A. & Gallet, C. (2001). What experimental protocol influences disparities between actual and hypothetical stated values? Environmental and Resource Economics, 20, 241-254.

List, J. A., Sadoff, S. & Wagner, M. (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. Experimental Economics, 14, 439-457.

Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies? Journal of Economic Surveys, 25, 363-370.

Loomis, J. & Eckstrand, E. (1998). Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: The case of the Mexican spotted owl. Ecological Economics, 27, 29-41.

McFadden, D. (2009). The human side of mechanism design: A tribute to Leo Hurwicz & Jean-Jacque Laffont. Review of Economic Design, 13, 77-100.

Meyer, B. D. (1995). Natural and quasi-experiments in economics. Journal of Business and Economic Statistics, 13, 151-161.

Montgomery, D. C. (2005). Design and analysis of experiments. New York, NY: John Wiley & Sons.

Murphy, J. J., Allen, P. G., Stevens, T. H. & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics, 30, 313-325.

Parsons, G. R. (1991). A note on choice of residential location in travel cost demand models. Land Economics, 67, 360-364.
Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: The discovered preference hypothesis. In K. J. Arrow, E. Colombatto, M. Perleman & C. Schmidt (Eds.), Rational foundations of economic behavior (pp. 225-250). New York, NY: St. Martin's.
Roth, A. E. (1995). Introduction to experimental economics. In J. H. Kagel & A. E. Roth (Eds.), The handbook of experimental economics (pp. 3-111). Princeton, NJ: Princeton University Press.
Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. Journal of Economic Theory, 10, 187-217.
Shogren, J. F., List, J. A. & Hayes, D. J. (2000). Preference learning in consecutive experimental auctions. American Journal of Agricultural Economics, 82, 1016-1021.
Slovic, P. (1995). The construction of preferences. American Psychologist, 50, 364-371.
Smith, V. L. (1964). Effect of market organization on competitive equilibrium. The Quarterly Journal of Economics, 78, 181-201.
Spybrook, J., Raudenbush, S. W., Liu, X. & Congdon, R. (2006). Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software. University of Michigan. Retrieved from: www.rmcs.buu.ac.th/statcenter/HLM.pdf.
Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.
Tversky, A. & Kahneman, D. (1981). The framing of decision and the psychology of choice. Science, 211, 453-458.
Vossler, C., Doyon, M. & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. American Economic Journal: Microeconomics, 4, 145-171.
Vossler, C. & Evans, M. F. (2009). Bridging the gap between the field and the lab: Environmental goods, policymaker input, and consequentiality. Journal of Environmental Economics and Management, 58, 338-345.
Vossler, C. & Watson, S. B. (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. Journal of Economic Behavior & Organization, 86, 137-147.
Whitehead, J. C., Dumas, C. F., Herstine, J., Hill, J. & Buerger, B. (2008). Valuing beach access and width with revealed and stated preference data. Marine Resource Economics, 23, 119-135.
Whitehead, J. C., Haab, T. C. & Huang, J.-C. (2000). Measuring recreation benefits of quality improvements with revealed and stated behavior data. Resource and Energy Economics, 22, 339-354.

Chapter 11
Benefit Transfer

Randall S. Rosenberger and John B. Loomis

Abstract Benefit transfer is a nonmarket valuation tool that is widely-used in a variety of decision contexts. Its primary role is deriving reliable estimates of value from prior research when new, original research is not feasible given time and resource constraints. This chapter begins by setting the development of benefit transfer in its historical context, which formally began in earnest in the early 1990's in response to an increased need for value measures in environmental and natural resource management and policy decisions. The two primary types of benefit transfer—value transfer and function transfer—are conceptually defined, including key steps when conducting them and examples of their empirical application. Sub-types of value transfers discussed include point estimate and measures of central tendency, and administratively-approved value transfers. Sub-types of function transfers discussed include benefit or demand function, and meta-regression analysis transfers. Reliability of benefit transfer is shown to be 45% transfer error for value transfers and 36% transfer error for function transfers. Criteria for minimizing transfer error rates in benefit transfers are provided to help guide practitioner decisions when using this tool.

Keywords Benefit transfer · Value transfer · Function transfer · Transfer error · Nonmarket valuation

Previous chapters of this book have described how to conduct an original nonmarket valuation study. However, original research may be time-consuming and expensive. What might an analyst do if a lack of time and/or funding prevents him or her from conducting an original study? For example, the U.S. Environmental Protection Agency is required to perform economic analyses on dozens of new

R.S. Rosenberger (✉)
Oregon State University, Corvallis, OR, USA
e-mail: R.Rosenberger@oregonstate.edu

J.B. Loomis
Colorado State University, Fort Collins, CO, USA
e-mail: John.Loomis@colostate.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_11

environmental regulations (e.g., water quality), but analysts are rarely given the time and budget to perform original studies (Griffiths et al. 2012). Thus, they must rely on existing data to compute the costs and benefits of programs. This chapter describes how existing data or summary statistics from previous research can be used in different decision contexts.

"Benefit transfer" is a colloquial term adopted by economists; it is defined as the use of existing data or information on nonmarket values in settings other than where they were originally collected. Benefit transfers are part of a broader family of economic data transfers. Sometimes an analyst may be interested in the price responsiveness of demand for certain goods or services. For example, what is the effect of a $3 fee increase on the demand for camping at a U.S. Forest Service campground, and what would be the total revenue generated from this fee increase? Here we are not necessarily interested in the total value of camping but in the change in use and potential revenue capture of a user fee program. Thus, information from nonmarket valuation studies can be used to inform policy and decision-making processes at various stages. It can be used in framing the policy context, evaluating policies (e.g., U.S. Environmental Protection Agency's assessment of the Clean Air Act [U.S. EPA 1997]), defining the extent of an affected market, prescreening of a natural resource damage assessment, and even determining whether original research is warranted.

While this chapter focuses on the use of benefit transfer to address valuation needs for nonmarket goods and services, the reader should not lose sight of its broader potential. The spirit of benefit transfer is that providing an estimate of the value of nonmarket resources using existing data may result in more balanced decision-making in situations where the direct monetary costs or opportunity costs of market commodities are known, but the nonmarket benefits are not (e.g., human health, public land recreation, ecosystem services, etc.). The purpose of this chapter is to provide guidelines to analysts to produce credible benefit transfer estimates. Before getting into the mechanics of benefit transfer, it is important to consider a historical context and a formal definition for the method.

11.1 A Historical Context

The first chapter of this book does a wonderful job of illustrating the need for value measures. Benefit transfer meets this need. Although economists have been estimating nonmarket values for more than half a century, the formal process of using benefit transfer to obtain estimates for nonmarket goods and services is only a few decades old. This short history reflects the age of nonmarket valuation.

The U.S. Army Corps of Engineers, the U.S. Bureau of Reclamation, and the U.S. Forest Service identified a need for estimates of recreation values for use in formal project evaluations and planning purposes. In 1973, the U.S. Water Resources Council published unit day value estimates of recreation activities for use in evaluating water-related projects. They updated these recreation value estimates in 1979 and 1983. In 1980, the U.S. Forest Service began publishing Resources Planning Act values (as per-person, per-activity-day estimates) for recreation (Rosenberger and Loomis 2001). Other Resource Planning Act values were published for timber, forage, minerals, and water. The U.S. Forest Service recreation estimates were driven by the Renewable Resources Planning Act of 1974, which required, among other things, a formal analysis of the costs and benefits associated with its programs.[1] Both the U.S. Water Resources Council's unit day values and the U.S. Forest Service's Resource Planning Act values were derived primarily from a combination of past empirical evidence, expert judgment, and political screening.

In the early 1980s, Freeman (1984) began the formal process of evaluating benefit transfer. He defined some specific conditions under which primary data could be transferable. In 1992, a section on benefit transfer was published in the journal Water Resources Research. Many of the top resource economists provided commentaries on the method in this special section. Collectively, the various articles suggested protocol, defined theory, identified needs, and presented new approaches.

Most benefit transfer exercises prior to the special section in Water Resources Research used a value transfer method that either transferred point estimates or measures of central tendency from original research or used administratively approved estimates. Loomis (1992) and Desvousges et al. (1992) proposed that more information, and therefore more robust benefit transfers, could be achieved with the transfer of entire demand (or benefit or willingness-to-pay) functions. Walsh et al. (1989, 1992) and Smith and Kaoru (1990) conducted meta-regression analyses of recreation values, potentially providing another approach to defining functions that could be used in benefit transfer.

The 1990s also were witness to the use of benefit transfers in the U.S. judicial system, primarily in natural resource damage assessments. While many of these early cases did not reach a jury verdict—instead concluding with negotiated settlements—the American Trading Transportation Company Inc. (ATTRANSCO) case was the first jury-defined verdict based on benefit transfer (see the text box that follows). The benefit transfer survived and informed the jury when setting the monetary value of recreation damages (Chapman and Hanemann 2001).

Since the early 1990s, new benefit transfer applications, tests, and methods were published in the literature. In 2000, the U.S. Environmental Protection Agency released its "Guidelines for Preparing Economic Analyses," in which steps for conducting benefit transfers were provided along with an approved estimate for the value of a statistical life (VSL) . These guidelines were updated in 2010 (U.S. EPA). In 2006, a special issue of the journal *Ecological Economics* reviewed state-of-the-art applications in benefit transfer. The focus of this special issue was primarily on function transfers, including value functions, meta-analysis functions, and structural preference functions. Conceptual issues in benefit transfers that go beyond a utility theoretical basis included the identification of various sources of

[1]The Government Performance and Results Act of 1993 superseded some previous federal legislation requiring formal cost-benefit analyses of federal programs.

transfer error (Rosenberger and Stanley 2006), temporal stability of values (Brouwer 2006), and impacts of primary valuation methods on value estimates (Johnston et al. 2006). Other contributions to the special issue reviewed applications of choice experiments (Morrison and Bergland 2006) and international transfers (Ready and Navrud 2006), as well as illustrating applications of geographic information systems (Troy and Wilson 2006) and spatial distributions (Bateman et al. 2006).

In 2000, the European Commission published its Water Framework Directive that seeks to improve surface and groundwater quality in Europe (European Commission 2005; WATECO 2004). As part of this directive, river basin management plans are required to be spatially integrative and measure the costs and benefits of these plans (Hanley and Black 2006). The nonmarket valuation requirements of this directive may necessitate wide use of benefit transfer (Hanley et al. 2006a) including transfers that rely on international databases (McComb et al. 2006).

In 2010, Johnston and Rosenberger provided a comprehensive review of benefit transfer applications and issues. Their discussion included the general issues of commodity definition consistency, primary study methodologies, spatial patterns, temporal trends, international benefit transfers, and site similarity.

In 2015, Johnston et al. expounded on their earlier work and published an edited book dedicated to benefit transfer. This book is a comprehensive guide written for researchers and practitioners, covering the theory, concepts, methods and applications of benefit transfer in environmental and natural resource economics.

Benefit Transfer in Action Problem

What is the value of lost recreation due to the American Trader oil spill in the Southern California area in 1990 (Chapman et al. 1998; Chapman and Hanemann 2001)? The primary needs are an estimate of lost recreation days and a value per recreation day. The following is part of the information provided by the plaintiffs in the trial People ex rel. Department of Fish and Game v. ATTRANSCO Inc. et al., Orange County, Ca., Superior Court Case No. 64 63 39, 1997.

Approach

Lost recreation days: Used extensive, historical to derive a function predicting number of recreation days lost due to beach closures and impeded access due to oil spill and mitigation efforts.

Value per recreation day: Existing literature was reviewed, resulting in Bell and Leeworthy's (1986) study of beach recreation in Florida being selected as the best available match for benefit transfer of beach-related recreation in Southern California.

Results

Original estimates as reported in Chapman and Hanemann (2001; 1990 $)

	Lost days	Value per day ($)	Total lost value ($)
Loss during beach closure period			
General beach recreation	454,280	13.19	5,991,953
Loss outside beach closure period			
General beach recreation	278,986	13.19	3,679,825
Total beach recreation loss			**$9,671,778**

This estimate was assumed to be conservative based on the income differential between Florida residents and Orange County, California residents. During the trial, expert economists for the plaintiffs and defendants testified about methodology, assumptions, and the results of different analyses to address areas of disagreement about data, countervailing factors, and modeling assumptions. In 1997, the jurors awarded the trustees $12.7 million in recreation damages, including estimates for beach, surfing, boating, fishing, and whale watching recreation, plus a civil liability of $5.3 million under the California Water Code, for a total award of $18 million.

11.2 Benefit Transfer Defined

When conducting an original nonmarket valuation study is not possible, values and other information from an existing study (or studies) can be used to estimate the value of the good or policy of current interest. Benefit transfer is the adaptation of existing value information to a new context. The context of the existing information is often referred to as the study site, and the benefit measure for the study site is defined as V_S.[2] The policy site, or the context for which information is needed, is defined as V_P. Ultimately, estimates of V_{Pj} for policy site j are derived from original research conducted at the study site i (V_{Si}). Study site values (V_{Si}) become transfer values (V_{Tj}) when applied to policy site j. However, as with any approximation, there is some error ε that needs to be minimized through the benefit transfer process.

$$V_{Si} = V_{Tj} + \varepsilon = V_{Pj}. \tag{11.1}$$

Original research provides content- and context-specific information regarding the policy site(s). This is because the target of original research is to address a specific need in a specific context. In the case of benefit transfer, the original

[2] V is used to denote value information or data and can consist of measures of benefits or costs, resource quantities or qualities, population characteristics, and other relevant information such as elasticities, dose-response effects, regression coefficients, and t-values.

research site and the policy site are often not the same (i.e., $i \neq j$). Therefore, the benefit transfer process must discuss the content- and context-relevant information of the study site and policy site in order for the user of the benefit transfer to have some idea of how well the study site matches the policy site. Ideally, the information transferred is relevant to the policy site context. Only in rare circumstances will the transferred information be specific to the policy site. Specificity would occur only if the study site and policy site were identical on all dimensions. In deciding whether the benefit transfer method is applicable, the analyst is trying to determine whether V_{Si} is similar enough in context to make a valid inference of V_{Pj}. This chapter discusses guidelines and approaches on how estimates of V_{Si} can be used to estimate V_{Pj} or the method of benefit transfer.

Now that a formal definition and a historical context of benefit transfer has been provided, the next section will clarify some of the terms used above, such as point estimate transfer, benefit function transfer, and meta-regression analysis transfer.

11.3 Modeling and Applying Benefit Transfer

Several benefit transfer methods have been developed to meet the needs for estimates of V_{Pj}, or the value at policy site j. These approaches are broadly classified as (1) value transfer and (2) function transfer. Value transfer involves the direct application of summary statistics from original research to a policy context. Function transfer involves the application of a statistical function that relates the summary statistics of original research to the specifics of the study site. In addition to providing the details of these approaches, the decision criteria an analyst could employ to choose the method to use are presented.

11.3.1 Value Transfer

Value transfer is the direct application of original research summary statistics (such as per-unit measures of willingness to pay [WTP], measures of elasticity, or other measures of marginal effects) to a policy site. There are essentially three approaches to conducting value transfers: (1) transfers of point estimates, (2) transfers of measures of central tendency, and (3) transfers of administratively approved estimates.

11.3.1.1 Point Estimate and Central Tendency Transfers

In some situations, a value transfer entails simply using a single study site estimate or an average of several estimates as the transfer value for a policy site. We call these types of value transfer a point estimate transfer or a measure of central

Table 11.1 Steps in conducting a point estimate value transfer

Step 1	Define the policy context (Q_{Pj}). This definition should include various characteristics of the policy site, what information is needed, and in what units
Step 2	Locate and gather original research outcomes (V_{Si}). Conduct a thorough literature review and obtain copies of potentially relevant publications
Step 3	Screen the original research studies for relevance ($Q_{Si} = Q_{Pj}$). How well does the original research context correspond to the policy context? Are the point estimates (V_{Si}) in the right units or can they be adjusted to the right units? What is the quality of the original research?
Step 4	Select a point estimate or range of point estimates (V_{Si}). This point estimate or range of point estimates should be selected on the best match between Q_{Pj} and Q_{Si}
Step 5	Transfer the point estimate or range of point estimates (V_{Tj}). Aggregate the point estimate to the policy site context by multiplying it by the total number of units, which provides a total value estimate for the good or service at the policy site

tendency (i.e., mean or median value) transfer, respectively. These transfers use measures of V_{Si}, given the context of study site i (Q_{Si}; e.g., location, population, resource changes at study site), to estimate the needed measure (V_{Pj}) for policy site j, given the context of the policy site (Q_{Pj}; e.g., location, population, resource changes at the policy site):

$$V_{Si}\big|Q_{Si} = V_{Tj} + \varepsilon = V_{Pj}\big|Q_{Pj}. \tag{11.2}$$

Point estimate transfer can be done using a single most similar site study value if there is one site that matches very closely or an average of site values if there is not a good single-site match. Given that there is unlikely to be a “near perfect” match between the study site and the policy site and if there are multiple values reported in the literature, then a range of estimates should be transferred to provide bounds on the value estimate at the policy site. In addition, it is recommended that confidence intervals be constructed around point estimate transfers when possible. This provides additional information regarding the precision of the study site measures. And furthermore, point estimates should be adjusted for differences in observable attributes of study and policy sites (e.g., adjusting the WTP value by measurable income differences; Ready and Navrud 2007).

Table 11.1 provides an overview of the steps in conducting a point estimate transfer. To illustrate this approach, a hypothetical application is provided. The Klamath River in Southwest Oregon is protected through designation as a national wild and scenic river by act of Congress in 1968. The river is 263 miles in length and is important to the region for its whitewater rafting, fishing, and natural beauty. However, the river also contains dams for the purposes of electricity generation and capturing water for agricultural irrigation. A coalition of tribes, conservationists, landowners and the local dam operator have come to an agreement that these dams should be removed. The hypothetical application of benefit transfer is to derive an estimate of the value of whitewater rafting for use in a benefit-cost analysis of the increased value of whitewater rafting due to dam removal.

The policy context in our example is defined as (1) the Klamath River in the Pacific Northwest, (2) use-value estimates for private (i.e., not commercially guided) whitewater rafting or kayaking, and (3) estimates of willingness to pay (i.e., consumer surplus) per person, per day. Given the defined policy context, a literature search was conducted. When conducting a literature search, the analyst should consider consulting experts, making inquiries on e-mail Listservs, and contacting relevant government agencies with management responsibilities for the resource of interest, in addition to keyword searches of electronic databases such as EconLit, AgEcon Search, and Google Scholar. The information the analyst seeks may not be provided in the peer-reviewed, published literature, so these contacts can help locate the "gray literature," such as theses and dissertations, unpublished working papers, conference proceedings, and agency reports. The gray literature may also help overcome the problem of selection bias and increase the number of estimates from which to choose (Stanley 2001; Rosenberger and Stanley 2006; Rosenberger and Johnston 2009).

The literature search could begin with keyword searches in the American Economic Association's EconLit database, Google Scholar, and Environment Canada's (1998) Environmental Values Reference Inventory Database. However, a related asset is the Recreation Use Values Database (Rosenberger 2011), which is a compilation of recreation use-value studies in the United States and Canada published through 2006. In addition to the literature referenced in the database, a working paper is added that estimated the value of whitewater rafting on a river in Colorado. Several studies located in the original search for constructing the database were discarded because they did not provide value estimates, which is the information sought after in this benefit transfer exercise.[3]

The initial search of the Recreation Use Values Database and other databases/tools resulted in 16 studies providing a total of 66 estimates of whitewater rafting or kayaking. Further evaluation of each of the studies and their estimates resulted in our discarding two studies and 14 estimates, primarily because they used nonstandard travel cost modeling procedures that lead to noncomparable outcomes. Table 11.2 lists some of the characteristics of the remaining 14 studies and 52 estimates of whitewater rafting or kayaking on 10 different rivers in the U.S. The two studies in Oregon for the Rogue River arguably provide the most similar site estimates of value given that the Klamath River, the policy context, is in the same general region as the Rogue River. In this case, the value per person, per day of private whitewater rafting or kayaking on the Rogue River is worth between $12 and $32 (in 2013 dollars). However, these estimates are based on a study that collected data nearly three decades ago, prior to dam removal on the Rogue River.

Alternatively, an analyst could select the most similar site based on measurable characteristics of each river. For the Klamath River, the average instream flow at its

[3]Studies that do not report any data or insufficiently report data may not be of use. Other factors can include a poor match between data needs for the policy site context (what is affected and how impacts are measured) and the context of the study site data. Boyle and Bergstrom (1992) describe how data may not be relevant for benefit transfers in general.

Table 11.2 Whitewater rafting/kayaking studies, United States (2013 $)

State/Region[a]	Rivers	Year studied	No. of studies	No. of estimates	Mean ($); (s.e.)	Range ($)
Maine	Dead	1994	1	4	41.84 (3.57)	31-47
Georgia, South Carolina	Chattooga	1979, 1993	2	8	242.64 (58.20)	21-457
North Carolina	Nantahala	1993	1	6	228.34 (24.84)	142-305
Idaho	Saint Joe, Salmon, Snake	1969, 1971, 1979, 2004	3	5	167.88 (81.29)	51-483
Utah	Colorado	1977	1	1	29.52 (–)	–
Colorado	Cache la Poudre	1978, 2010	2	3	77.18 (19.36)	39-99
Arizona	Colorado	1985, 1998	2	15	204.84 (30.73)	12-380
Wyoming	Snake	2004	1	2	219.61 (120.14)	99-340
California	Tuolumne	1982	1	2	108.00 (20.39)	88-128
Oregon	Rogue	1984	2	6	20.33 (2.82)	12-32
Mountain Region	5	–	8	26	177.40 (25.69)	12-483
Pacific Region	2	–	3	8	42.25 (15.00)	12-128
West Region	7	–	11	34	145.60 (22.20)	12-483
Total	10	–	14[b]	52	$162.09 (18.49)	$12-483

[a]Regional estimates are supersets of states' estimates
[b]The number of studies does not add up due to two studies evaluating rivers in multiple states

mouth is 17,010 cubic feet per second (cfs); it has a change in elevation of 4,090 feet; and it is 263 miles in length. The most similar river based on instream flow at its mouth is the Colorado River, with potential transfer estimates ranging from $12 to $379 per person, per day for whitewater recreation. The most similar river based on elevation change is the Nantahala River in North Carolina, with potential transfer estimates of $190 to $456 per person, per day for whitewater recreation. And the river most similar in length is the Rogue River, with potential estimates ranging from $12 to $32 per person, per day for whitewater recreation. None of the rivers in the Recreation Use Values Database best match the Klamath River policy site on more than one measurable characteristic. However, if studies are filtered by the region they focus on, then the Rogue River remains the most similar site.

Table 11.3 Whitewater rafting/kayaking studies, United States (2013 $), 20% trimmed mean

Region	No. of studies	No. of estimates	Mean ($); (s.e.)	Range ($)
Mountain Region	6	17	140.15 (20.39)	39-294
Pacific Region	2	3	82.53 (28.06)	32-128
West Region	8	20	131.51 (18.23)	32-294
Total	**10**	**32**	**$142.63 (15.42)**	**$31-294**

However, the two characteristics of region and river length are not sufficient in themselves to declare the Rogue River as the most similar site. In addition, comparison with the broader literature does lead to some concern about relying solely on the Rogue River value estimates given they are at the lower bound of the distribution across all value estimates for whitewater recreation. This example suggests that selecting the most similar site is not necessarily a simple matter given that different sites may match on different characteristics.

A measure of central tendency transfer involves using a mean, median, or other measure of central tendency based on all or a subset of original research outcomes. Continuing with the previous example of estimating the value of a private whitewater rafting or kayaking day on the Klamath River, note that the first three steps are the same as the point estimate transfer. The next step is to calculate a measure of central tendency for all of the estimates. In this case, the information from all studies is used, thus minimizing concerns about point estimate transfers discussed previously. The average of all 52 estimates is $162 per person, per day (Table 11.2), an order of magnitude larger than the site-specific estimates discussed previously. The range is from $12 to $483. If the average value is restricted to the Western Census Region due to concerns about comparability of whitewater experiences in the Eastern U.S., then the average of the remaining 34 estimates is $146, still with the same range in estimates. However, if the region is further restricted to the Pacific Census Division, the average of the remaining eight estimates—which is the Rogue River estimates combined with two estimates for the Tuolumne River in California—is $42, with a range of $12 to $128 per person, per day.

Furthermore, analysts should be concerned about the vagaries of random sampling and modeling assumptions on the range of estimates from the literature. Objectively trimming the tails of the distribution of estimates from the literature reduces the effect of very low and very high estimates on the calculated average value. For example, Table 11.3 reports summary statistics from trimming 20% of the observations (10% from each tail of the distribution). This technique reduces the number of studies from 14 to 10 and the number of estimates from 52 to 32. Specifically, five of the six Rogue River specific estimates are removed because they fall in the 10% trimmed from the lower tail, and the remaining estimate defines the limit of the lower tail. The mean value from the 32 estimates is $143 per person, per day, which is not statistically different from the original average value for the entire data. However, the Pacific Census Division estimate is now $82 per person, per day, due to the increased influence of the Tuolumne River study in California. The same arguments that were made regarding the effect of the point estimate benefit transfer measure on the policy process can be made here. That is, if the estimate, when combined with other values associated with dam removal, is substantially greater than or smaller than the costs, then the benefit transfer exercise has added information to the policy process. If a more precise estimate could change the project or policy decision, then original research is warranted.

11.3.1.2 Administratively Approved Estimate Transfer

Administratively approved estimate transfer is arguably the simplest approach to benefit transfer if these measures are available and match your data needs. In the U. S., federal agencies commonly use administratively approved estimates in assessing management and policy actions. The U.S. Forest Service has used Resource Planning Act values since 1980 (Rosenberger and Loomis 2001). These Resource Planning Act values are provided for groups of outdoor recreation activities and Forest Service regions of the country. Similarly, the U.S. Bureau of Reclamation and U.S. Army Corps of Engineers have relied on the U.S. Water Resources Council's unit day values of recreation use for decades (U.S. Water Resources Council 1973, 1979, 1983) and continue to do so today. The most significant—and controversial—administratively approved estimate is in the form of the value of statistical life (VSL). The VSL is a statistical measure of income or wealth that would be traded off for a small change in the probability or risk of death; conversely, it measures the tradeoff between money and life. The U.S. Environmental Protection Agency has published its recommended "default central 'value of statistical life' (VSL) of $7.9 million (in 2008 dollars) to value reduced mortality for all programs and policies" (U.S. EPA 2010, pp. 7-8).

Administratively approved estimates are derived from empirical evidence in the literature, expert judgment, and political screening. There are two main issues associated with using administratively approved estimates. First, the criteria used in the political screening process are unknown. This process may ignore some empirical evidence or use arbitrary adjustment factors. The second issue is that administratively approved estimates are only updated occasionally. Therefore, estimates may not reflect the latest empirical evidence. One distinct advantage of using administratively approved estimates for agency purposes is that the estimates have survived the political screening process.

The administratively approved transfer can be defined as

$$V_{SA} = V_{Tj} + \varepsilon = V_{Pj}, \tag{11.3}$$

where V_{SA} is the administratively approved measure that is transferred to provide the value at policy site j (V_{Pj}). Table 11.4 outlines the steps involved in conducting an administratively approved estimate transfer. Analysts should not use these measures solely on the basis that the U.S. Forest Service, U.S. Army Corp of Engineers, or U.S. Environmental Protection Agency uses them. Administratively approved estimates are developed to address certain agency needs, which may or

Table 11.4 Steps in conducting an administratively approved estimate transfer

Step 1	Define the policy context (Q_{Pj}). This definition should include various characteristics of the policy site, what information is needed, and in what units
Step 2	Obtain administratively approved estimate (V_{SA}). These estimates are typically published by an agency. Check with the relevant agency's policy or research division
Step 3	Transfer the administratively approved estimate (V_{Tj}). Aggregate the estimate to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site

may not match well with the analyst's needs. The analyst should understand how and why these values were developed and then determine whether they meet his or her needs.

U.S. Environmental Protection Agency's uses of its administratively approved VSL central estimate include the Clean Air Interstate Rule, the Nonroad Diesel Rule, and the Stage 2 Disinfectants and Disinfection Byproducts Rule (see U.S. EPA, 2010, for links to these rules and other information regarding VSLs). While the VSL estimate is approved for use in important policy developments, it is based on 26 studies conducted between 1974 and 1991, primarily using hedonic wage methods across a variety of populations and types of death. The Agency is aware of limitations and implicit errors associated with this VSL estimate, including measurement error issues in original studies and the age of studies behind the estimate; however, until it updates the studies used or develops other methods, this estimate continues to be recommended and used in practice.

11.3.2 Function Transfer

Function transfers are more technically oriented than value transfers. They involve the transfer of functions or statistical models that define relationships between value estimates and study site characteristics. Some of these models were introduced previously in this book. Function transfers can be categorized as demand (or benefit or WTP) functions[4] or meta-regression analysis functions.

Function transfers are generally considered to perform better than value transfers, an issue investigated in more detail at the end of this chapter. This increased accuracy is because function transfers may be tailored to fit some of the characteristics of the policy site. Value transfers, on the other hand, are invariant to most differences between the study site and the policy site. However, as previously noted, value transfers can be adjusted for some attributes (e.g., income) that differ between the study site and policy site.

[4]Other functions include dose-response or production functions, especially prevalent in the health sciences literature.

11.3.2.1 Demand or Benefit Function Transfer

Demand or benefit function transfers are based on the premise that the study site estimate for site i (V_{Si}) is a function of characteristics of the study site context (Q_{Si}; e.g., location, physical features, and climate) and other explanatory variables (Z_{Si}; e.g., sociodemographics, attitudes):

$$V_{Si} = Vf(Q_{Si}, Z_{Si}). \tag{11.4}$$

Other chapters in this book provided reasons why and how to estimate WTP and demand functions using a variety of nonmarket valuation tools. This additional information may be used to take advantage of these relationships when conducting benefit transfer. Rather than relying solely on value transfers, precision may be gained from incorporating these relationships in benefit transfer. A value transfer requires a strong similarity between study sites and policy sites, which may not always be the case. The invariance of value transfer measures to other relevant characteristics of a policy site may make these transfers insensitive or less robust to significant differences between the study site and the policy site. Therefore, the precision of benefit transfer can be increased if value estimates are adapted via a function to fit the specifics of a policy site, under conditions that the underlying structural relationships (e.g., preferences, behaviors) are stable across sites.

The beginning steps to conducting a demand or benefit function transfer are the same as for a value transfer with the exception that additional information is required from publications. Some form of a function that models the statistical relationships between the summary measures of interest and characteristics of the original research effort—including characteristics of the study site and the study population—must be reported if a study is to be used in a benefit function transfer. The analyst will ultimately adapt this function to specific characteristics of the policy site, thereby predicting values for the policy site. A near-perfect match between the study site and policy site is not required because the analyst can potentially compensate for these differences in function transfers.

The demand or benefit function transfer can be defined as

$$Vf_S\left(Q_{S|Pj}, Z_{S|Pj}\right) = V_{Tj} + \varepsilon = V_{Pj}. \tag{11.5}$$

The policy site measure (V_{Tj}) is derived from the study site function $\left(Vf_S\right)$ adjusted to the characteristics of the policy site ($Q_{S|P}$ and $Z_{S|P}$). This is why, in Step 4 of Table 11.5, summary data are gathered on the policy site for as many of the independent, or explanatory, variables in the model as possible. This information is used to tailor or adapt the study site function to the policy site context.

The following is an example of a benefit function transfer. This example is simplistic,[5] but it illustrates several issues: (1) application of benefit function

[5]Another reason this example is simplified is that it deals with a benefit function, which is a direct estimation method. As such, it directly models the relationship between WTP and independent

Table 11.5 Steps in conducting a demand or benefit function transfer

Step 1	Define the policy context (V_{Pj}). This definition should include various characteristics of the policy site (Z_{Pj}), what information is needed, and in what units
Step 2	Locate and gather original research demand or benefit functions (Vf_S). Conduct a thorough literature review and obtain copies of potentially relevant publications
Step 3	Screen the original research studies for relevance ($Q_{Si} = Q_{Pj}$). How well does the original research context correspond to the policy context? What is the quality of the original research? And most importantly, is a demand or benefit function (Vf_S) provided?
Step 4	The demand or benefit function (Vf_S) provided by original research has several independent or explanatory variables associated with it. Gather summary data on the policy site (Z_{Pj}) for as many of the variables in the model as possible
Step 5	Predict the policy site benefit estimate (V_{Tj}) by multiplying the summary statistics reflecting the policy site by the regression coefficients in the transfer function ($Q_{S\|P}$ and $Z_{S\|P}$). This results in a tailored estimate for the policy site
Step 6	Aggregate the tailored estimate to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site

transfers, (2) effect of multisite data modeling on transfer accuracy, and (3) validity testing between value and function transfer approaches. Assume that the value of interest is for improving groundwater quality used for drinking to a very safe level in a small Northeastern town. A literature search identified several groundwater quality improvement studies. From among all of the studies identified, the following study seems to provide the most relevant benefit function.

A case study by VandenBerg et al. (2001) estimates the benefits of improving groundwater quality used for drinking to a "very safe" level in 12 towns in New York, Massachusetts, and Pennsylvania. The study also tests the relative accuracy of two benefit transfer approaches. The authors used a contingent valuation survey. Mean WTP per household, per year was calculated for each of the 12 towns using the survey responses. This mean WTP is treated as the benchmark or known estimate (V_{Pj}) for each town j to which the transferred estimate (V_{Ti}) is compared. They used the estimate derived for each of the other 11 towns (study sites) as possible point measures to transfer to the 12th town (the policy site).

To perform the benefit function transfer, a protocol first used by Loomis (1992) is employed whereby all of the survey data except for one town were pooled and a WTP equation was estimated. The independent variables of this function were then set at the levels for the "*n*th," or excluded, town in order to predict mean WTP per household at this excluded town. In all, 12 benefit function models were estimated.

(Footnote 5 continued)

variables. Other models, such as demand models, may not be as easily adjusted or may not be amenable to adjustment depending on how the models are developed, including functional form (Adamowicz et al. 1989).

VandenBerg et al. (2001) report a benefit function model based on pooling the data for all 12 towns, which will be used later in this example.

A sufficient number of explanatory variables that account for relevant differences in the site attributes (e.g., size, quality) between the study site and policy site, as well as key differences in demographics (e.g., income, average age, education) should be included. In the VandenBerg et al. (2001) example, explanatory variables included demographics such as education and income. These types of data, which are available from secondary sources, allow the analyst to control for differences in demographics between the study site and the policy site. However, if these data are not part of the demand or benefit function at the study site, then adjusting for these particular expected differences between the two sites is not feasible. Benefit transfer estimates using an incomplete benefit transfer function adjustment are likely, on average, to be more accurate than value transfers with point estimates from study sites that are most similar to the policy site.

VandenBerg et al. (2001) also included a series of responses to risk perceptions, perceptions regarding current safety of water, and perceived likelihood of future contamination. These variables were statistically significant and contributed to the explanatory power of their models. However, they make real-world benefit transfer difficult because the analyst would need to know the perception variables for the policy site. Typically, these perception variables are not available from secondary sources and would require a survey of households at the policy site. Benefit function transfer reduces the amount of information needed for valuation and, thus, only a simple survey would need to be conducted in order to collect the desired information. In real-world applications, the analyst may simply need to use the original study site perceptions, noting the assumption being made that these are the same between the study and policy site populations. Ideally the analyst would like to have explanatory variables in the benefit transfer function that are available at both the study site and the policy site. Thus, when the analyst is choosing among possible benefit functions, he or she should keep in mind whether information is available to support a more detailed benefit function transfer versus a simpler but less data-demanding benefit function.

VandenBerg et al. (2001) provide a benefit function (mean WTP per household, per year) based on pooling all the data for the 12 towns (Table 11.6). This model contains several dummy variables that would enable adjustment for differences between the study and policy populations. No variables that identify physical characteristics of the sites are included in the model. Most of the variables would require conducting a survey at the policy site in order to determine how its population differs from the populations used to estimate the model. However, they do provide summary statistics for most of the variables for each town in the data set; typically, this is what the benefit transfer analyst would rely on in performing a benefit function transfer. The last two columns in Table 11.6 show the adjustments to this function to reflect characteristics of one of the towns.

In order to tailor or adapt VandenBerg et al.'s (2001) benefit function to the policy site, the regression coefficients from the model are multiplied by measures of the variables for the policy site to derive the partial WTP associated with each of the

Table 11.6 Ordinary least squares regression model for groundwater quality protection, Dependent Variable = mean willingness to pay per household, per year, $N = 667$

Variable	Regression coefficient	Policy site measure	Partial WTP
Constant	−29.68	1	−29.68
Perception of contamination (0, 1 = no experience)	−23.48	1	−23.48
Perception of contamination (0, 1 = don't know)	−26.96	0.24	−6.47
Likelihood of future contamination (0, 1 = likely or very likely)	17.51	0.5	8.76
Likelihood of future contamination (0, 1 = not sure)	9.41	0.5	4.70
Interest in community water issues (0, 1 = mild to no interest)	−20.66	0.5	−10.33
Interest in community water issues (0, 1 = interested)	−11.15	0.5	−5.58
Perceived water quality (0, 1 = unsafe or somewhat safe)	29.92	0.5	14.96
Perceived water quality (0, 1 = safe)	21.07	0.5	10.54
College degree (0, 1 = has college degree)	−17.51	0.24	−4.20
Some college (0, 1 = has some college but no degree)	−15.72	0.24	−3.77
Average risk perception (3-question composite; 1 = safe to 5 = unsafe)	9.91	4	39.64
Number of perceived potential contamination sources	2.56	3.17	8.12
Trust in government and organizations (9-question composite; 1 = do not trust to 3 = trust)	15.67	1.91	29.93
Household income ($/year)	0.0008	45,500	36.40
Total mean WTP = $\sum$(column 2 $\times$ column 3) =			**$69.54**

Note Adjusted $R^2 = 0.15$
Adapted from VandenBerg et al. (2001)

variables. See VandenBerg et al. (2001) for a full account of the different variables and site measures. Values for the independent variables are for the Horsham, Pennsylvania, policy site.

The first variable in the model, the constant or intercept term, is transferred in full to the policy site. The next eight variables measure various factors associated with perceived contamination and water safety and community contamination issues. Given this type of information is likely not available for a policy site, associated variables would be adjusted based on reported values for the study site. For more details regarding these variables, see VandenBerg et al. (2001).

Assigning values for the policy site to the remaining variables in the model is relatively straightforward. The college education variables do not perfectly match with the summary statistic reported for the policy site. The policy site reports that 48% of the town is college-educated based on the survey responses. What proportion of this 48% has a college degree versus some college is unknown. Because the regression coefficients for the education dummy variables are similar, one way to capture the effect of education is to split this proportion in half, which transfers 24% of each education variable for a total 48% effect of college education.

For the policy site, VandenBerg et al. (2001) report that the mean average risk perception is 4, the mean number of perceived potential contamination sources is 3.17, the mean trust is 1.91, and mean household income per year is $45,500 for Horsham. Thus, multiplying the respective coefficients by these reported levels for the policy site provides the partial WTP effects of these variables.

Summing all of the partial effects listed in the last column of Table 11.6 results in $69.54 in total mean WTP per household, per year for the policy site town of Horsham, Pennsylvania. The actual mean WTP per household, per year for Horsham is $67.45, based on site-specific data. The benefit function transfer estimate is well within a 95% confidence interval for the actual value for the site and has a percent transfer error of about 3%.

11.3.2.2 Meta-Regression Analysis Function Transfer

Another function transfer approach that is gaining application is the meta-regression analysis function transfer. Demand or benefit function transfers rely on statistical relationships defined for certain variables based on a single study. Meta-regression analysis summarizes and synthesizes outcomes from several studies. There are essentially two approaches to meta-regression analysis: (1) pooling the actual data from multiple studies and (2) using summary statistics, such as value estimates, from multiple studies. The latter is more prevalent and the focus of this section.

Meta-regression analysis overcomes some of the issues related to demand or benefit function transfers. Namely, it is possible to statistically explain the variation found across empirical studies (as found in the whitewater rafting or kayaking example). Some of this variation in value estimates may be due to identifiable characteristics among the different studies themselves, such as valuation method, survey mode, geographic location, and so forth. These characteristics are not explanatory variables in the original studies because they define the context of original research and are, therefore, constant in original research. Meta-regression analysis may be able to discern the individual study effects some of these variables may have on estimated values.

Meta-regression analysis has traditionally been concerned with understanding the influence of methodological and study-specific factors on research outcomes and providing summaries and syntheses of past research (Stanley and Jarrell 1989). The first two meta-analyses on environmental and natural resource economic studies were by Smith and Kaoru (1990) on travel cost studies of recreation benefits and by Walsh et al. (1989, 1992) on outdoor recreation benefit studies. Since then, meta-analysis has become a rapidly expanding method, although not always aimed at benefit transfer applications. Nelson and Kennedy (2009) identify and evaluate more than 130 distinct applications of meta-analysis in environmental economics, with the majority conducted since 2003.

The dependent variable in a meta-regression analysis is a summary statistic from each individual study, such as a value estimate, elasticity, or other measure. The independent or explanatory variables are characteristics of the model, survey design, and data of the original studies. The interstudy variation in research outcomes may be explained by modeling the characteristics that are typically held constant within an individual study, such as valuation methodology, survey mode, time, and physical attributes of the study site.

A basic premise of meta-regression analysis is the existence of an underlying valuation function, of which original research studies are independent random draws. This premise is likely false for many literatures (Rosenberger and Johnston 2009). The draws are not random because a reason exists for conducting original research on some sites and not others (selection bias). Peer-review screening for statistically significant results in journals (publication bias) is also an issue. The draws are probably not independent due to multiple estimates from single studies or

Table 11.7 Steps in conducting a meta-analysis function transfer

Step 1	Define the policy context (Q_{Pj}). This definition should include various characteristics of the policy site, what information is needed, and in what units
Step 2	Develop a standard database structure. Conduct a thorough literature review and obtain copies of potentially relevant publications. Develop a master coding strategy that allows for consistently coding as much information as possible regarding each study. This information includes the dependent (V_{Si}) and independent variables in the original analysis, methodological (M_{Si}) and other study characteristics (Z_{Si}), source of the study, and authors of the study
Step 3	Before coding the individual studies, screen the original research studies for relevance ($Q_{Si} = Q_{Pj}$). Reduce the literature search outcomes to include those studies containing relevant empirical estimates, tests, or findings
Step 4	Choose and reduce the summary statistic (V_{Si}) to a common metric. The summary statistic would be the primary information needed for the policy site. Reduction to a common metric may include reducing all empirical estimates to the same unit (e.g., annual basis). This summary statistic will serve as the dependent variable in the meta-analysis regression
Step 5	Choose the independent variables (Z, M). These variables are those characteristics of the individual studies that are hypothesized to be important or consequential to differences in the summary statistics
Step 6	Conduct the meta-regression analysis ($Vf_S(Q, Z, M)$). The summary statistic will serve as the dependent variable, and the independent variables will serve as the explanatory variables. The purpose of meta-regression analysis is to explain the variation in the dependent variable across studies. Standard econometric issues are relevant here
Step 7	Gather summary data for the policy site (Z_{Pj}). The meta-regression analysis model has several associated independent variables. Gather summary data on the policy site for as many of the variables in the model as possible
Step 8	Predict the policy site summary statistics (V_{Tj}) from the meta-regression model (Vf_S) by multiplying the summary statistics reflecting the policy site (Q_P, Z_P, M_P) by the regression coefficients ($Vf_S(Q_{S\|P}, Z_{S\|P}, M_{S\|P})$) in the transfer function. This results in a tailored estimate for the policy site (V_{Tj})
Step 9	Aggregate the tailored estimate to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site

Adapted from Stanley (2001)

Table 11.8 Summary statistics for whitewater rafting/kayaking metadata

Variable	Description	Mean values	
		Full data	20% trimmed
South	1 = Southern Census Region; 0 = otherwise[a]	0.269	0.250
West	1 = Western Census Region; 0 = otherwise[a]	0.654	0.625
Site quality	1 = site quality is rated high by author; 0 = otherwise	0.673	0.625
Sample frame	1 = sample is drawn from onsite visitors; 0 = otherwise	0.519	0.594
Valuation method	1 = travel cost model; 0 = otherwise	0.519	0.406
Trip type	1 = private trip; 0 = otherwise	0.558	0.500
Average flow	Average river flow in cubic feet per second (cfs)	14,204	13,975
Change in elevation	Elevation change: Source to mouth of river in miles	6,576	6,374
River length	Length of river in miles	606	578
Wild and scenic	1 = portion of river is designated wild and scenic; 0 = otherwise	0.500	0.438
No. of observations	Number of estimates of recreation value	52	32

[a]Omitted category is Northeastern Census Region (plus no observations in Midwestern Census Region)

from researchers who work closely together. There is also the potential for autocorrelation due to learning effects and improvements in methodology over time.

The steps to conducting a meta-regression analysis (Table 11.7) are adapted from Stanley (2001). An application of meta-regression analysis is illustrated by evaluating the studies and estimates included in Table 11.2 and the reduced set in Table 11.3. Again, the policy target is to derive an estimate of private whitewater rafting or kayaking on the Klamath River.

Table 11.8 provides variable descriptions and summary statistics for the full and 20% trimmed metadata. Specific characteristics gleaned from each study included the census region of the study, the quality of the site as stated by the authors of each study, the sample frame and valuation method used by the authors, and the type of recreation trip (i.e., private versus commercially guided trips). Information for each study location was augmented with proxy measures for river quality and characteristics, including the average river flow, overall change in elevation of the river from its source to its mouth, the overall length of each river, and whether some portion of the river has been federally designated as a national wild and scenic river. While these proxies for whitewater qualities are crude, they are systematically related to variations in value estimates reported in the literature. Better proxies of river quality might be obtained through the use of geographic information systems data (e.g., see Ghermandi et al., 2010), which may reduce the inherent measurement error when deriving proxy measures through data augmentation.

Next a regression analysis is conducted where the valuation estimates adjusted for units and inflation are the dependent variable and the variables described in Table 11.8 are the independent variables. The goal of the regression model is to explain as much of the variation in the dependent variable across the studies as possible using the independent variables. Correcting the model for common statistical issues (e.g., multicollinearity and heteroskedasticity) is relevant here (Nelson and Kennedy 2009). After testing for and affirming a normal distribution for the dependent variable, an ordinary least squares model is estimated. Several of the independent variables cannot be included in the regression because insufficient information is provided by individual studies. For example, most of the studies did not report any summary statistics regarding the characteristics of their sample population, such as mean income, age, education, etc. These variables may be important to explaining some of the variation in the value estimates, but they cannot be tested when information on them is not provided in the original studies.

Table 11.9 provides the meta-regression analysis results for the full metadata and the 20% trimmed metadata. Also included in the table is an example application of these meta-regressions as benefit function transfer, which will be discussed below. For now, focus on the two columns reporting coefficient estimates. The dependent variable is the natural log of the value per person, per day, as reported by each study. Overall, these simple meta-regressions fit the data very well, with R^2 values of 0.65 and 0.79 for the full metadata and 20% trimmed metadata samples, respectively. The regression results show that rivers studied in the Southern Census Region have statistically higher value estimates than rivers studied elsewhere; average river flow and overall change in elevation for each river are statistically and positively related to value estimates; and river length is statistically and negatively related to value estimates for the full metadata meta-regression analysis. For the 20% trimmed metadata, the regression results are consistent in the direction of measured effects (i.e., positive or negative) with the full metadata regression (the exception is valuation method, which is insignificant in both models). However, the statistical significance of some of the estimated coefficients has changed. Up to this point, the meta-regression analyses provided statistical summaries or quantitative reviews of the literature. Now the meta-analysis regressions are used as benefit transfer functions.[6]

The meta-regression analysis transfer function is defined as

$$Vf_S\left(Q_{S|Pj}, Z_{S|Pj}, M_{S|Pj}\right) = V_{Tj} + \varepsilon = V_{Pj}. \tag{11.6}$$

Equation (11.6) states that the value for policy site j (V_{Pj}) is a function of data included in or distinguishable from each study site i. The other variables can be quantity/quality variables (Q); socio-demographic variables (e.g., income, age, and

[6]The potential use of meta-regression analysis in defining benefit transfer functions is like the holy grail of benefit transfer: developing a function that can be used to estimate different types of values for different policy contexts. That is, even in conditions where no point estimates or demand functions are reported in the literature, a meta-regression analysis function may be able to provide such estimates or functions.

Table 11.9 Meta-regression analysis and function predictions, whitewater rafting/kayaking on the klamath river

Variable	Full data MRA[a]			20% trimmed data MRA[b]		
	Coefficient	Policy value[c]	Increment[d]	Coefficient	Policy value[c]	Increment[d]
Constant	3.30478*	1	3.30	3.52250*	1	3.52
South	1.23337*	0	0	1.17058*	0	0
West	−1.05715	1	−1.06	−0.22583	1	−0.22
Site quality	0.57336	1	0.57	0.67515*	1	0.68
Sample frame	0.77846	0.52	0.40	0.48632	0.59	0.29
Valuation method	−0.86976	0.52	−0.45	0.08210	0.41	0.03
Trip type	−0.59010	1	−0.59	−0.33194*	1	−0.33
Average flow	0.00003*	17010	0.22	0.00001*	17010	0.09
Change in elevation	0.00057*	4090	3.03	0.00018	4090	0.98
River length	−0.00317*	263	−0.68	−0.00091	263	−0.20
Wild and scenic	−1.24168	1	−1.24	−0.80212*	1	−0.80
Root MSE[e]	0.72056			0.40129		
R^2	0.6504			0.7866		
Predicted value[f]			$25.94			$53.94
95% confidence interval[g]			$17-40			$36-82

*Significant at the 0.10 significance level. Dependent variable is natural log of consumer surplus per person, per day (2013 $)

[a]Ordinary least squares cluster robust regressions where clusters are defined by individual studies (No. clusters = 14)

[b]Ordinary least squares cluster robust regressions where clusters are defined by individual studies (No. clusters = 10)

[c]Policy values are set to best match policy site conditions except for methodology variables, which are set at their mean values for the metadata

[d]Incremental effects are coefficients × policy values. The sum of increments = natural log of consumer surplus per person, per day (2013 $)

[e]Root mean square error (MSE) = $\sqrt{\sigma^2}$

[f]Predicted value calculated as exp(σ^2/2) * exp(predicted natural log of consumer surplus per person per day)

[g]95% confidence intervals calculated as predicted value ± 1.96 + se (θ) [se(θ) is standard error of the fitted regression]

education) and site characteristics (e.g., presence of water, geographic location, and land type; *Z*); and methodological variables (e.g., valuation method, modeling format, and functional form; *M*) for each study (*i*). The application of this model to benefit transfer is similar to adjusting the benefit function discussed in the previous section. Value estimates tailored to a policy site may be obtained by adjusting the meta-regression analysis function to specific characteristics of that site.

To recap previous benefit transfer applications, a literature search was conducted that identified two documents reporting value estimates seemingly specific to the policy site. However, an analyst may be concerned about the age of these studies and/or the implicit modeling assumptions used by the primary study authors. This led to a broader literature search on whitewater rafting or kayaking in the United States. Expanding on the literature search criteria resulted in 14 studies reporting a total of 52 estimates. An average value from the literature could be calculated. The effect of very low or very high estimates on calculated average values is explored through trimming the tails of the distribution from this literature. Now, meta-regression analysis functions are used that control for systematic relationships among the data to predict estimates of value for the policy context.

Table 11.9 illustrates how the meta-regression functions can be used to predict value estimates. The variables are set to levels that match the policy site (see Policy Value columns in Table 11.9), including the partial effects of the region, site quality, trip type, and wild and scenic river status by setting these parameter values at 1. The partial effect of the South Census Region is negated in the model by setting its parameter value to 0. There is no a priori reason to judge the sample frame or valuation method, so these parameter values are set at their mean values in each metadata sample (this approach includes the average partial effect of methodology variables as included in the metadata). The partial effects of river flow, change in elevation, and river length are set to match the policy site's characteristics. The predicted value for whitewater rafting or kayaking on the Klamath River is \$26 per person, per day, with a 95% confidence interval of \$17 to \$40 for the full metadata function. The 20% trimmed metadata function results in a prediction of \$54 per person, per day, with a 95% confidence interval of \$36 to \$82. Based on overlapping confidence intervals, these two predictions are not statistically different from each other; however, the overlap is modest and the absolute magnitudes of the average predicted values would lead to different aggregate measures of recreation benefits.

Just as an analyst may be concerned that the most similar site point estimate transfer values are too low relative to the rest of the empirical literature, the average value transfers for the entire data, regional data, or trimmed data may be too large (Tables 11.2 and 11.3). The meta-regression benefit transfer functions' predictions may be relatively more defensible given they not only rely on the broader valuation literature but enable the adjustment of values to the policy site context based on measurable differences among study sites. The prediction from the full metadata regression prediction has a confidence interval that includes estimates from two of the most similar studies, while the trimmed metadata regression prediction is more than the most similar study's estimates. Further augmenting the metadata may lead to better (i.e., more defensible) predictions, although the need for precise, valid measures may only be derived through conducting an original study, that is, if the analyst has the resources available to make such a choice. If not, then benefit transfer may be the best and only option for deriving value estimates.

11.3.2.3 Structural Benefit Function Transfers

One of the limitations of the benefit transfer methods discussed in this chapter is the general lack of a micro-level theoretical foundation that clearly links consumer utility functions to the benefit transfer process (Johnston and Rosenberger 2010). This formal modeling of the utility functions for benefit transfer applies to both value and function transfers, which were previously discussed.

Bergstrom and Taylor (2006) tackle this issue by distinguishing between strong structural utility theoretic models, weak structural utility theoretic models, and nonstructural utility theoretic models. Strong models explicitly derive a benefit transfer function or meta-analysis equation from a utility function. At the other end of the spectrum, nonstructural models do not attempt to explicitly link the benefit transfer function or meta-analysis to an explicit utility function. Rather, they merely attempt to specify a benefit transfer function or meta-analysis that would be consistent with economic theory. As Bergstrom and Taylor (2006) note, strong structural utility theoretical models are preferred, and weak structural utility models are acceptable on theoretical grounds. They state that nonstructural utility theoretical models are not suggested for performing meta-analysis benefit transfer. Obviously, this recommendation must be balanced with whether time and budget are available to assemble data needed for a strong structural benefit transfer or whether existing “off-the-shelf,” weak structural utility theoretic-based meta-analyses will have to be used or not. If these weak structural meta-analyses models are used due to expediency, this limitation should be pointed out.

Smith et al. (2002) proposed an alternative benefit transfer method that they call structural benefit transfers or preference calibration transfers. In structural benefit transfers, the analyst specifies a utility function that describes an individual’s choices over a set of market and nonmarket goods or services, assuming a standard budget-constrained utility maximization problem (Smith et al. 2006). An analytical expression is derived that links available benefit measures to the assumed utility function, which defines the theoretical foundation for transfers. Calibration of benefit measures and other pertinent information is conducted to ensure the consistency of measures across each study. While these models impose theoretical consistency on the use of prior information, they are also limited in that they require a strong a priori assumption regarding the underlying structural model. Furthermore, the structural benefit transfer method is more complex than the weak structural utility theoretic models discussed previously, which may be one reason it has not been widely adopted or applied (Johnston and Rosenberger 2010). Nonetheless, there is wide agreement that strong structural utility theoretic models have many potential benefits over weak structural utility theoretic models (McConnell 1992; Boyle et al. 1994; Bergstrom and Taylor 2006).

Table 11.10 Summary of absolute percentage transfer error (|PTE|) by research studies[a]

Transfer type	Median |PTE|	Mean |PTE|	Std. error of mean |PTE|	*N*
Value	45	140	10.6	1,792
Function	36	65	4.0	756

[a]See http://recvaluation.forestry.oregonstate.edu/sites/default/files/PTE_Summary.pdf for a table listing individual studies, percentage transfer errors, and other pertinent information on specific benefit transfer error studies

11.4 Accuracy of Benefit Transfers: How Good Are Benefit Transfers?

Assessing the precision and accuracy of a particular benefit transfer is usually impossible because the actual value for a policy site (V_{Pj}) is unknown; otherwise, there would be no need for benefit transfer. If the best approximation of the actual value for a policy site is not known, then how does the analyst know how close the benefit transfer (V_{Tj}) is to the actual value (V_{Pj})? It is like playing pin the tail on the donkey without the donkey. How can one know how close he or she is to the target when no target exists? Therefore, in order to assess the validity and reliability of benefit transfer, the target's value must be known. Validity tests of benefit transfer include access to some original research; that is, where $V_S = V_P$ when $i = j$. Based on outcomes from benefit transfer validity and reliability studies, indicators emerge as to when and why some transfers are better than others (i.e., when ε in $V_{Tj} + \varepsilon = V_{Pj}$ is small), which are discussed later in this chapter.

Several studies have evaluated the validity and size of transfer error ε associated with benefit transfers in varying contexts (Bergstrom and DeCivita 1999; Brouwer and Spaninks 1999; Morrison and Bergland 2006; Rosenberger and Stanley 2006; Johnston and Rosenberger 2010; Rosenberger 2015). The magnitude of acceptable transfer error may vary based on the specific context for the transfer. For example, lower transfer errors (i.e., more precision) may be needed for the calculation of compensatory amounts in negotiated settlements and litigation than situations requiring broad benefit-cost analyses as information gathering or screening of projects and policies (Navrud and Pruckner 1997).

There are generally two categories of benefit transfer errors: measurement error and generalization error (Rosenberger and Stanley 2006). Measurement errors are associated with the original research that provides value estimates used in benefit transfer. While these errors are embedded in the information to be transferred, an analyst may minimize these errors by evaluating the quality of original research—a transfer can only be as good as the information on which it relies. Generalization, or transfer, errors are errors associated with the transfer process itself, including poor correspondence between the study and policy sites and the type of transfer method employed. Matching the contexts of the study and policy sites helps minimize transfer error (Boyle et al. 2009, 2010).

Table 11.10 provides a summary of the various studies that directly evaluated benefit transfer accuracy. In particular, these studies measured percentage transfer error (PTE) as follows:

$$PTE = \left[\left(\left(V_{Tj} - V_{Pj}\right)/V_{Pj}\right)\right] \times 100, \tag{11.7}$$

where V_{Tj} is the transferred estimate and V_{Pj} is the known policy site estimate. Summary statistics reported for each benefit transfer study include the number, median, mean, and range of absolute values of PTE (|PTE|). Each study is further classified by resource type, primary valuation method, and benefit transfer category (i.e., value or function transfer). This information will help the analyst identify specific types of benefit transfer tests he or she may want to investigate further given the context of the analyst's own benefit transfer.

There are 38 studies behind the summary statistics reported in Table 11.10, with almost half of them (N = 18) evaluating value and function transfers within the same context. There are 1,792 estimates of PTE for value transfers and 756 estimates of PTE for function transfers. Value transfers have a higher range in possible PTE estimates than function transfers, although both ranges are large. Value transfers have a mean |PTE| of 140% and median of 45%. Function transfers have a mean |PTE| of 65% and a median of 36%. The differences in mean and median |PTE| for value versus function transfers are statistically significant at the 99% confidence level, supporting the conclusion that function transfers outperform value transfers in general (Johnston and Rosenberger 2010; Kaul et al. 2013; Rosenberger 2015).

Based on median |PTE|, benefit transfers generally perform well, although if improperly done, they can result in substantial transfer error. It is often suggested that following some best practice guidelines for benefit transfer will reduce transfer error, providing the analyst with confidence in specific applications (Boyle et al. 2009, 2010; Johnston and Rosenberger 2010). A quantitative assessment of factors associated with varying levels of PTE is provided by Kaul et al. (2013) and Rosenberger (2015), suggesting a few patterns in the benefit transfer literature. Errors are generally found to be smaller in cases where sites and populations are most similar (Rosenberger and Phipps 2007). Studies that illustrate the importance of site correspondence include Loomis (1992), Piper and Martin (2001), Rosenberger and Loomis (2000), VandenBerg et al. (2001), Barton (2002), Morrison et al. (2002), Morrison and Bennett (2004, 2006), Johnston (2007), and Colombo and Hanley (2008).

Function transfers are shown to generally be better than value transfers. Intrastudy comparisons of value versus function transfers show that function transfers result in a lower mean and median |PTE| than value transfers the majority of the time[7] (e.g., Loomis 1992; Parsons and Kealy 1994; Bergland et al. 2002; Bowker et al. 1997; Kirchhoff et al. 1997; VandenBerg et al. 2001; Groothius 2005; Kristofersson and Navrud 2007; Matthews et al. 2009; Boyle et al. 2010;

[7]The range of PTE estimates from the literature are provided in an appendix located at http://recvaluation.forestry.oregonstate.edu/sites/default/files/PTE_Summary.pdf.

Rosenberger 2015). In some applications, given both the study site and policy site measures are provided in the same original research, many estimates of value are compared—not because they should be, but because enough information is provided to do so. These naïve transfers illustrate the risk of high transfer errors if inappropriate transfers are conducted. For example, Lindhjem and Navrud (2008) compare several types of transfers and analysts' assumptions. Their results show that PTE magnitude and range are reduced when screening for best fit or using a meta-analysis function to predict policy site values. And finally, interpretations of PTE as validity indicators are weak because study site values themselves are estimated with error, leaving real transfer errors largely unknown (Rosenberger and Stanley 2006; Boyle et al. 2010).

The literature reports other types of validity tests, including a difference in means test and a difference in model coefficients test (Rosenberger 2015). These are tests wherein value and model coefficient estimates and their standard errors are compared between study sites and policy site applications. In general, these tests reject the null hypotheses that the means and coefficients are equal the majority of the time, and they have a weak positive correlation with PTE measures. While these particular validity tests suggest benefit transfers are not valid, they often fail to recognize the context of benefit transfers and acceptable levels of accuracy. Furthermore, these validity tests show a counterintuitive result in that less efficient statistical estimates (i.e., larger standard errors of estimates) have a greater probability of failing to reject equality compared with more efficient estimates (i.e., smaller standard errors of estimates), implying greater transferability (Kristofersson and Navrud 2005). Nonetheless, standard hypothesis tests remain the norm in the benefit transfer literature (Lindhjem and Navrud 2008).

Alternative types of validity tests have been proposed, but not widely adopted (Lerman 1981; Desvousges et al. 1998; Spash and Vatn 2006; Lindhjem and Navrud 2008). For example, there is literature that applies equivalence testing within benefit transfer (Muthke and Holm-Mueller 2004; Kristofersson and Navrud 2005; Hanley et al. 2006a; Johnston 2007; Johnston and Duke 2008). Equivalence testing changes the burden of proof in traditional hypothesis testing by reversing the null and alternative hypotheses (i.e., estimates are assumed different unless tests show the difference is smaller than a specified tolerance limit and probability value). In benefit transfer applications, the tolerance limit is specified as the maximum acceptable level of transfer error for which transfer and policy estimates are considered equivalent (e.g., Kristofersson and Navrud, 2007, and Johnston and Duke, 2008, use tolerance limits of 40%, while Baskaran et al., 2010, use 50 and 80%). These tolerance limits may be set based on the context of the benefit transfer exercises, as noted previously (Navrud and Pruckner 1997).

11.5 Criteria for Choosing Among Benefit Transfer Approaches

An evaluation of published benefit transfer studies suggests that there is no single best method. How does the analyst choose the method that is the best fit for his or her application? Ultimately, the best choice might be to minimize expected transfer error. While the choices may not be straightforward or simple, five criteria are identified that are useful for guiding these decisions. A point estimate transfer may be preferred when the available study site estimates closely match the policy site on (1) the good being valued, including quantity and quality, activity type, resource attributes (e.g., water clarity), or species of interest; (2) the geographic area being evaluated; (3) the affected population and its characteristics; (4) the welfare measure (e.g., property rights assignments, WTP); and (5) the valuation methods used in the study site application are conceptually, theoretically, and empirically sound.

The analyst may not be able to match all five criteria between available studies and defined policy needs. As the number of matches dwindles, the analyst may be best served to move to other methods that rely on functional relationships among the values and underlying data. If a valuation function is available in the set of candidate studies, then a function transfer approach may be more accurate than point estimate transfers (at least based on the summary of validity results found in Table 11.10). In part, the advantage of valuation function transfers is when there is not a good match between the study site and policy site on some of the criteria, but information is available that enables adjustments to be made.

However, there may be times when a valuation function cannot be found that matches some of the five factors, and the additional information needed for adjusting estimates is not available. In this case, if several studies are available that collectively match the five criteria, then an average value or some other measure of central tendency might be transferred with greater defensibility than relying on a valuation function from any one of the studies.

If at this point the analyst is left with only two to three studies that reasonably match his or her policy needs, then two other options are available. The most defensible choice would be to consider developing a meta-analysis regression transfer function, assuming the literature is robust enough for statistical modeling. As noted previously, meta-regression functions provide a means for the analyst to capture and control for many of the differences between the policy site and the available literature. Any resulting benefit estimates should be treated as a "generic value" that is primarily indicative of the range of likely value estimates.

Another, less defensible, option is to use a "back-of-the-envelope" method based on transferring an administratively approved value. Choice of this method is often based on the analyst not having training or experience in nonmarket valuation. While this method is used by agencies and other groups, its use is not recommended until all other options have been exhausted.

In some circumstances, there may be no clear "superior" approach for conducting the benefit transfer. In other words, no one method can, even collectively, satisfy all five of the factors. In this case, the analyst should evaluate all potential methods, given information constraints, and apply the ones that seem plausible in order to provide the decision-maker with a range of estimates that reflect the uncertainty in benefit transfer estimates. A range of estimates of benefits may be sufficient for the purpose of some economic analysis, and it avoids providing the decision-maker with a false sense of precision that providing a single estimate might convey.

At what point might the analyst simply indicate that no defensible estimate of value can be derived from benefit transfer? It depends on the purpose of the benefit transfer value. If the benefit transfer value is merely being included so the analysis acknowledges that there are economic benefits received from the nonmarketed resource, then even unit day values may make this point better than omission of any value at all. However, if the benefit-cost decision is likely to hinge on the values calculated from one of these weaker benefit transfer methods, it may be necessary for the analyst to present the limited valuation evidence to the decision-maker, along with a recommendation that even a simple original primary valuation study is likely to yield more accurate results than reliance on a mismatched benefit transfer estimate (possibly guided by the decision method presented in Allen and Loomis, 2008). Ultimately, it is up to the decision-makers to make the call because they will have to defend their policy or project decision.

11.6 Conclusions

This chapter was intended to expose the reader to the world of benefit transfer. The method of benefit transfer is described along with example applications. The reader's understanding should now be sufficient to critique benefit transfer or even conduct his or her own benefit transfer.

In the end, the analyst must decide whether he or she can perform a defensible benefit transfer that will improve decision-making by an approximation of the value of some nonmarket resource (e.g., recreation, water quality, etc.) versus decision-making with no information on the economic values of affected nonmarket resources. When there is uncertainty about which benefit transfer method to apply or what assumptions must be made to apply a particular benefit transfer method, the analyst should present a range of value estimates. In some cases, such a range of economic values of these nonmarket resources may be sufficient to determine whether this range is above or below the cost of a particular policy or project. In other cases, providing decision-makers with a rough idea of the magnitude of nonmarket values helps change their perceptions of the relative values of natural resources or ecosystem services flowing from their area (Ervin et al. 2012). Timely valuation information is often more useful for decision-making than no estimate whatsoever, but the analyst must guard against the tendency to go out on the "benefit transfer limb" of feeling compelled to provide some estimate, regardless of its accuracy. "Incredible" estimates will undermine the overall credibility of benefit transfer in particular and nonmarket valuation in general.

References

Adamowicz, W. L., Fletcher, J. J. & Graham-Tomasi, T. (1989). Functional form and the statistical properties of welfare measures. American Journal of Agricultural Economics, 71, 414-421.

Allen, B. P. & Loomis, J. B. (2008). The decision to use benefit transfer or conduct original valuation research for benefit-cost and policy analysis. Contemporary Economic Policy, 26, 1-12.

Barton, D. N. (2002). The transferability of benefit transfer: Contingent valuation of water quality improvements in Costa Rica. Ecological Economics, 42, 147-164.

Baskaran, R., Cullen, R. & Colombo, S. (2010). Testing different types of benefit transfer in valuation of ecosystem services: New Zealand winegrowing case studies. Ecological Economics, 69, 1010-1022.

Bateman, I. J., Day, B. H., Georgiou, S. & Lake, I. (2006). The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP. Ecological Economics, 60, 450-460.

Bell, F. W. & Leeworthy, V. R. (1986). An economic analysis on the importance of saltwater beaches in Florida. Florida Sea Grant Report SGR-82. Gainesville: University of Florida.

Bergland, O, Magnussen, K. & Navrud, S. (2002). Benefit transfer: Testing for accuracy and reliability. In R. J. G. M. Florax, P. Nijkamp & K. G. Willis (Eds.), Comparative environmental economic assessment (pp. 117-132). Cheltenham, United Kingdom: Edward Elgar.

Bergstrom, J. C. & DeCivita, P. (1999). Status of benefit transfer in the United States and Canada: A review. Canadian Journal of Agricultural Economics, 47 (1), 79-87.

Bergstrom, J. C. & Taylor, L. O. (2006). Using meta-analysis for benefits transfer: Theory and practice. Ecological Economics, 60, 351-360.

Bowker, J. M., English, D. B. K. & Bergstrom, J. C. (1997). Benefit transfer and count data travel cost models: An application and test of a varying parameter approach with guided whitewater rafting. Working Paper FS 97-03. Athens: University of Georgia.

Boyle, K. J. & Bergstrom, J. C. (1992). Benefit transfer studies: Myths, pragmatism, and idealism. Water Resources Research, 28, 657-663.

Boyle, K. J., Kuminoff, N. V., Parmeter, C. F. & Pope, J. C. (2009). Necessary conditions for valid benefit transfers. American Journal of Agricultural Economics, 91, 1328-1334.

Boyle, K. J., Kuminoff, N. V., Parmeter, C. F. & Pope, J. C. (2010). The benefit-transfer challenges. Annual Review of Resource Economics, 2, 161-182.

Boyle, K. J., Poe, G. L. & Bergstrom, J. C. (1994). What do we know about groundwater values? Preliminary implications from a meta analysis of. American Journal of Agricultural Economics, 76, 1055-1061.

Brouwer, R. (2006). Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event. Ecological Economics, 60, 399-406.

Brouwer, R. & Spaninks, F. A. (1999). The validity of environmental benefits transfer: Further empirical testing. Environmental and Resource Economics, 14, 95-117.

Chapman, D. J. & Hanemann, W. M. (2001). Environmental damages in court: The American Trader case. In A. Heyes (Ed.), The law and economics of the environment (pp. 319-367). Cheltenham, United Kingdom: Edward Elgar.

Chapman, D. J., Hanemann, W. M. & Ruud, P. (1998). The American Trader oil spill: A view from the beaches. Association of Environmental and Resource Economists (AERE) Newsletter, 18 (2), 12-25.

Colombo, S. & Hanley, N. (2008). How can we reduce the errors from benefits transfer? An investigation using the choice experiment method. Land Economics, 84, 128-147.

Desvousges, W. H., Johnson, F. R. & Banzhaf, H. S. (1998). Environmental policy analysis with limited information: Principles and applications of the transfer method. Cheltenham, United Kingdom: Edward Elgar.

Desvousges, W. H., Naughton, M. C. & Parsons, G. R. (1992). Benefit transfer: Conceptual problems in estimating water quality benefits using existing studies. Water Resources Research, 28, 675-683.

Environment Canada. (1998). Environmental Valuation Reference Inventory homepage. Retrieved from www.evri.ca.

Ervin, D., Larsen, G. & Shinn, C. (2012). Simple ecosystem service valuation can impact national forest management. Association of Environmental and Resource Economist (AERE) Newsletter, 32 (1), 17-22.

European Commission. (2005). Overall approach to the classification of ecological status and ecological potential. Common Implementation Strategy for the Water Framework Directive, Guidance Document No. 13. Luxembourg: European Commission.

Freeman, A. M. III. (1984). On the tactics of benefit estimation under Executive Order 12291. In V. K. Smith (Ed.), Environmental policy under Reagan's executive order: The role of benefit-cost analysis (pp. 167-186). Chapel Hill: University of North Carolina Press.

Ghermandi, A., van den Bergh, J. C. J. M., Brander, L. M., de Groot, H. L. F. & Nunes, P. A. L. D. (2010). Values of natural and human-made wetlands: A meta-analysis. Water Resources Research, 46.
DOI 10.1029/2010WR009071.

Griffiths, C., Klemick, H., Massey, M., Moore, C., Newbold, S., Simpson, D., et al. (2012). U.S. Environmental Protection Agency valuation of surface water quality improvements. Review of Environmental Economics and Policy, 6, 130-146.

Groothius, P. A. (2005). Benefit transfer: A comparison of approaches. Growth and Change, 36, 551-564.

Hanley, N. & Black, A. (2006). Cost benefit analysis and the Water Framework Directive in Scotland. Integrated Environmental Assessment and Management, 2, 156-165.

Hanley, N., Colombo, S., Tinch, D., Black, A., & Aftab, A. (2006). Estimating the benefits of water quality improvements under the Water Framework Directive: Are benefits transferable? European Review of Agricultural Economics, 33, 391-413.

Hanley, N., Wright, R. E. & Alvarez-Farizo, B. (2006). Estimating the economic value of improvements in river ecology using choice experiments: An application to the water framework directive. Journal of Environmental Management, 78, 183-193.

Johnston, R. J. (2007). Choice experiments, site similarity and benefits transfer. Environmental and Resource Economics, 38, 331-351.

Johnston, R. J., Besedin, E. Y. & Ranson, M. H. (2006). Characterizing the effects of valuation methodology in function-based benefits transfer. Ecological Economics, 60, 407-419.

Johnston, R. J. & Duke, J. M. (2008). Benefit transfer equivalence tests with non-normal distributions. Environmental and Resource Economics, 41, 1-23.

Johnston, R. J., Rolfe, J., Rosenberger, R. S. & Brouwer, R. (editors) (2015). Benefit transfer of environmental and resource values: A handbook for researchers and practitioners. Dordrecht, The Netherlands: Springer. 606p.

Johnston, R. J. & Rosenberger, R. S. (2010). Methods, trends and controversies in contemporary benefit transfer. Journal of Economic Surveys, 24, 479-510.

Kaul, S., Boyle, K. J., Kuminoff, N. V., Parmeter, C. F. & Pope, J. C. (2013). What can we learn from benefit transfer errors? Evidence from 20 years of research on convergent validity. Journal of Environmental Economics and Management, 66, 90-104.

Kirchhoff, S., Colby, B. G. & LaFrance, J. T. (1997). Evaluating the performance of benefit transfer: An empirical inquiry. Journal of Environmental Economics and Management, 33, 75-93.

Kristofersson, D. & Navrud, S. (2005). Validity tests of benefit transfer – Are we performing the wrong tests? Environmental and Resource Economics, 30, 279-286.

Kristofersson, D. & Navrud, S. (2007). Can use and non-use values be transferred across countries? In S. Navrud & R. Ready (Eds.), Environmental values transfer: Issues and methods (pp. 207-225). Dordrecht, The Netherlands: Springer.

Lerman, S. R. (1981). A comment on interspatial, intraspatial, and temporal transferability. In P. R. Stopher, A. H. Meyburg & W. Borg (Eds.), New horizons in travel-behavior research (pp. 628-632). Lexington, MA: Lexington Books.
Lindhjem, H. & Navrud, S. (2008). How reliable are meta-analyses for international benefit transfers? Ecological Economics, 66, 425-435.
Loomis, J. B. (1992). The evolution of a more rigorous approach to benefit transfer: Benefit function transfer. Water Resources Research, 28, 701-705.
Matthews, D. I., Hutchinson, W. G. & Scarpa, R. (2009). Testing the stability of the benefit transfer function for discrete choice contingent valuation data. Journal of Forest Economics, 15, 131-146.
McComb, G., Lantz, V., Nash, K. & Rittmaster, R. (2006). International valuation databases: Overview, methods and operational issues. Ecological Economics, 60, 461-472.
McConnell, K. E. (1992). Model building and judgment: Implications for benefit transfers with travel cost models. Water Resources Research, 28, 695-700.
Morrison, M. & Bennett, J. (2004). Valuing New South Wales rivers for use in benefit transfer. Australian Journal of Agricultural and Resource Economics, 48, 591-611.
Morrison, M. & Bennett, J. (2006). Valuing New South Wales rivers for use in benefit transfer. In J. Rolfe & J. Bennett (Eds.), Choice modelling and the transfer of environmental values (pp. 71-96). Cheltenham, United Kingdom: Edward Elgar.
Morrison, M. & Bergland, O. (2006). Prospects for the use of choice modelling for benefit transfer. Ecological Economics, 60, 420-428.
Morrison, M., Bennett, J., Blamey, R. & Louviere, J. (2002). Choice modeling and tests of benefit transfer. American Journal of Agricultural Economics, 84, 161-170.
Muthke, T. & Holm-Mueller, K. (2004). National and international benefit transfer with a rigorous test procedure. Environmental and Resource Economics, 29, 323-336.
Navrud, S. & Pruckner, G. J. (1997). Environmental valuation – To use or not to use? A comparative study of the United States and Europe. Environmental and Resource Economics, 10, 1-26.
Nelson, J. P. & Kennedy, P. E. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. Environmental and Resource Economics, 42, 345-377.
Parsons, G. R. & Kealy, M. J. (1994). Benefits transfer in a random utility model of recreation. Water Resources Research, 30, 2477-2484.
Piper, S. & Martin, W. E. (2001). Evaluating the accuracy of the benefit transfer method: A rural water supply application in the USA. Journal of Environmental Management, 63, 223-235.
Ready, R. & Navrud, S. (2006). International benefit transfer: Methods and validity tests. Ecological Economics, 60, 429-434.
Ready, R. & Navrud, S. (2007). Morbidity value transfer. In S. Navrud & R. Ready (Eds.), Environmental values transfer: Issues and methods (pp. 77-88). Dordrecht, The Netherlands: Springer.
Rosenberger, R. S. (2011). Recreation Use Values Database. Corvallis: Oregon State University. Retrieved Feb. 15, 2013, from http://recvaluation.forestry.oregonstate.edu/sites/default/files/RECREATION_USE_VALUES_DATABASE_%20SUMMARY.pdf.
Rosenberger, R. (2015). Chapter 14: Benefit transfer validity, reliability and error. In R. J. Johnston, J. Rolfe, R. S. Rosenberger & R. Brouwer (eds.), Benefit transfer of environmental and resource values: A handbook for researchers and practitioners (307-326). Dordrecht, The Netherlands: Springer.
Rosenberger, R. S. & Johnston, R. J. (2009). Selection effects in meta-analysis and benefit transfer: Avoiding unintended consequences. Land Economics, 85, 410-428.
Rosenberger, R. S. & Loomis, J. B. (2000). Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database. Water Resources Research, 36, 1097-1107.
Rosenberger, R. S. & Loomis, J. B. (2001). Benefit transfer of outdoor recreation use values: A technical document supporting the Forest Service Strategic Plan (2000 Revision). General

Technical Report RMRS-GTR-72. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Rosenberger, R. & Phipps, T. (2007). Correspondence and convergence in benefit transfer accuracy: Meta-analytic review of the literature. In S. Navrud & R. Ready (Eds.), Environmental value transfer: Issues and methods (pp. 23-43). Dordrecht, The Netherlands: Springer.

Rosenberger, R. S. & Stanley, T. D. (2006). Measurement, generalization, and publication: Sources of error in benefit transfers and their management. Ecological Economics, 60, 372-378.

Smith, V. K. & Kaoru, Y. (1990). Signals or noise? Explaining the variation in recreation benefit estimates. American Journal of Agricultural Economics, 72, 419-433.

Smith, V. K., Pattanayak, S. K. & van Houtven, G. (2006). Structural benefit transfer: An example using VSL estimates. Ecological Economics, 60, 361-371.

Smith, V. K., van Houtven, G. & Pattanayak, S. K. (2002). Benefit transfer via preference calibration: "Prudential algebra" for policy. Land Economics, 78, 132-152.

Spash, C. L. & Vatn, A. (2006). Transferring environmental value estimates: Issues and alternatives. Ecological Economics, 60, 379-388.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. Journal of Economic Perspectives, 15 (3), 131-150.

Stanley, T. D. & Jarrell, S. B. (1989). Meta-regression analysis: A quantitative method of literature surveys. Journal of Economic Surveys, 3, 161-170.

Troy, A. & Wilson, M. A. (2006). Mapping ecosystem services: Practical challenges and opportunities in linking GIS and value transfer. Ecological Economics, 60, 435-449.

U.S. EPA (U.S. Environmental Protection Agency). (1997). The benefits and costs of the Clean Air Act, 1970 to 1990. Washington, DC: U.S. EPA. Retrieved from www.epa.gov/economics.

U.S. EPA (U.S. Environmental Protection Agency) National Center for Environmental Economics, Office of Policy. (2010). Guidelines for preparing economic analyses. Washington, DC: U.S. EPA. Retrieved from yosemite.epa.gov/ee/epa/eerm.nsf/vwAN/EE-0568-50.pdf/$file/EE-0568-50.pdf.

U.S. Water Resources Council. (1973). Principles, standards, and procedures for water and related land resource planning. Federal Register, 38, 24778-24945.

U.S. Water Resources Council. (1979). Procedures for evaluation of National Economic Development (NED) benefits and costs in water resources planning (Level c). Federal Register, 44, 72892-72976.

U.S. Water Resources Council. (1983). Economic and environmental principles and guidelines for water and related land resources implementation studies. Washington, DC: U.S. Government Printing Office.

VandenBerg, T. P., Poe, G. L. & Powell, J. R. (2001). Assessing the accuracy of benefits transfers: Evidence from a multi-site contingent valuation study of groundwater quality. In J. C. Bergstrom, K. J. Boyle & G. L. Poe (Eds.), The economic value of water quality (pp. 100-120). Cheltenham, United Kingdom: Edward Elgar.

Walsh, R. G., Johnson, D. M. & McKean, J. R. (1989). Issues in nonmarket valuation and policy application: A retrospective glance. Western Journal of Agricultural Economics, 14, 178-188.

Walsh, R. G., Johnson, D. M. & McKean, J. R. (1992). Benefit transfer of outdoor recreation demand studies: 1968-1988. Water Resources Research, 28, 707-713.

WATECO. (2004). Economics and the environment: The implementation challenge of the Water Framework Directive: A Guidance Document. Luxembourg: European Commission.

Chapter 12
Reliability and Validity in Nonmarket Valuation

Richard C. Bishop and Kevin J. Boyle

Abstract A central goal of nonmarket valuation studies is to provide accurate value estimates. We suggest a systematic, comprehensive approach to accuracy assessment using the twin concepts of reliability and validity. Reliability has to do with variance; validity has to do with bias. If procedures applied in a valuation study are produce erratic results, accuracy suffers even if those procedures are unbiased. And, even if the procedures in question are reliable, they will be less useful if the procedures applied produce large biases in the estimates. We adapt the general concepts of reliability and validity to apply to nonmarket valuation studies. As in many other disciplines, the concept to be measured, the "true values," is unobservable. Hence, criteria must be developed to serve as indicators of accuracy. Reliability is typically observed using the estimated standard error of the mean from repeated estimates of value. Validity is assessed using the concepts of content validity, construct validity, and criterion validity; what we refer to as the "three Cs" of validity. After fleshing out these concept, we illustrate how they can be applied using two case studies. Contingent valuation serves as an example of stated preference methods. The travel cost method provides a case study of revealed preference methods.

Keywords Nonmarket valuation · Reliability · Test-retest · Validity · Content validity · Construct validity · Criterion validity · Contingent valuation · Travel cost

When researchers apply the methods described in this book, they strive to estimate values that are accurate. But how are they to judge the accuracy of their results? Or, if reviewers are assigned to evaluate the accuracy of value estimates coming from a

R.C. Bishop (✉)
University of Wisconsin-Madison, Madison, WI, USA
e-mail: rcbishop@wisc.edu

K.J. Boyle
Virginia Tech, Blacksburg, VA, USA
e-mail: kjboyle@vt.edu

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8_12

nonmarket valuation study, what criteria should they bring to bear? This chapter addresses issues of accuracy based on the concepts of reliability and validity.

If you Google "reliability and validity," you will quickly learn two things. First, the joint concepts have been applied to address accuracy issues in many different branches of science. And second, the different sciences have adapted these two concepts to meet their specific needs. How the concepts are applied in epidemiology may be different from, say, social psychology. This chapter will explain how reliability and validity can be used to assess the accuracy of nonmarket value estimates.

Nearly everything discussed in this chapter on the topic of accuracy has been addressed somewhere in the nonmarket valuation literature, but the literature has examined accuracy in a piecemeal fashion. What is sought here is a framework to support systematic and comprehensive assessments of the accuracy of nonmarket value estimates.

On an intuitive level, reliability and validity can be understood in terms of shooting arrows at a target. The reliability of the archer is revealed by whether the arrows are tightly grouped or scattered about on the target. Validity is revealed by whether the arrows are centered around the bull's-eye or around some other point on the target. Succinctly stated, reliability is concerned with error dispersion; validity addresses systematic error or bias.

To start the adaptation of the concepts of reliability and validity to nonmarket valuation, assume that a set of valuation procedures can be applied repeatedly in the same research setting. Each application is like a shot at the target. Reliability is defined based on the standard error of the value estimates from these repeated applications. The smaller the standard error, the more reliable the resulting value estimates are judged to be. Validity is defined based on differences between values in the individual applications and the "true value." For example, maximum willingness to pay, as defined in Chap. 2, is represented by the bull's-eye on a target, and each estimate would be a shot at the target. In this metaphor, bias is the distance of the average from all shots from the center of the target—true willingness to pay. In theory, validity assessment focuses on the expected value of differences between observed value estimates and latent, unobserved willingness to pay. The smaller the expected value of the differences, the more valid are the value estimates and the estimation method itself.

In the real world, researchers only have one or very few opportunities to apply nonmarket valuation procedures in a particular setting. Similarly, they only get one or at most a few opportunities to assess the reliability and validity of value estimates from individual studies. Nevertheless, the more that prior studies of a method's reliability and validity support the specific procedures that have implemented, the more confident one can be that the results of an individual study are accurate (i.e., come close to the true value). This chapter discusses assessment of the validity and reliability of individual studies.

In nonmarket valuation, validity and reliability are assessed at the study level using summary statistics, not at the level of individual observations. Further, to the extent that an individual study is deemed to be reliable and valid, this outcome supports the reliability and validity of the valuation method itself and the procedures used to implement the method. Recursively, reliability and validity research helps identify valuation methods and implementation procedures that can be used later on to implement a reliable and valid study. The prior chapters in this book focus on the use of such procedures when applying nonmarket valuation methods.

Middle sections of this chapter present a framework for considering the evidence on reliability and validity. While practical reliability assessment is straightforward, validity assessment is more complicated and difficult and will require a larger discussion. For reasons that will be explained momentarily, while the true value is a useful theoretical construct, it cannot be observed in the real world. Hence, indirect evidence of validity must be considered in applied validity assessments.

The different kinds of evidence of validity are organized around the "three C's" of validity: content validity, construct validity, and criterion validity. Content validity assessment involves consideration of the research procedures applied in the application. Construct validity assessment considers how the results from a nonmarket valuation study stand up to hypothesis tests based on prior expectations such as those formed by consistency with economic theory. Criterion validity assessment considers how well results from the procedures employed in the research compare to results from procedures that are widely considered to have a high degree of validity. Final judgments about the validity of nonmarket values depend on the weight of evidence derived from these three parts of validity assessment.

Final sections of the chapter illustrate how the framework can be applied by considering the reliability and validity of two nonmarket valuation methods. Contingent valuation will serve as a case study of a stated preference method, and the travel cost method will illustrate how the framework applies to a revealed preference method.

12.1 Introduction to Reliability and Validity

Let's begin with the concept of true value, as the term applies in nonmarket valuation. The monetary value of something is measured in terms of either willingness to pay (WTP) or willingness to accept compensation (WTA). Recalling Chap. 2, consider a program that would improve environmental quality. If the improvement is not made, an individual consumer will enjoy some level of utility, say v^0. If the improvement in environmental quality were accomplished, it would put the person on a higher level of utility, $v^1 > v^0$. The true value, WTP_i (in this case, compensating surplus), for this change is defined as the maximum income that consumer i could give up once the environmental change takes place and still have the same utility as before the change, v^0. This is equivalent to saying that WTP is the maximum amount of money a person could pay and be exactly indifferent between realizing and not realizing the environmental change. This indifference-producing amount of money is what researchers normally try to measure in nonmarket valuation studies, which is the true WTP_i for the improvement and is defined by

$$v_i\left(P^0, Q^0, y_i\right) = v_i\left(P^0, Q^1, y_i - \mathrm{WTP}_i\right), \tag{12.1}$$

where $v_i(.)$ is the indirect utility function of the ith consumer, P^0 is a vector of prices assumed here to be constant, Q is a vector of environmental quality attributes where at least one element increases from Q^0 to Q^1, and y_i is the ith consumer's income. WTP is subscripted by i to denote that willingness to pay can vary across individuals.

To economize on words, this chapter will be developed assuming WTP is the goal of measurement. But this is done with the understanding that, depending on circumstances, the goal could be measurement of WTA and that everything said about reliability and validity would continue to apply. Further, the concepts discussed herein are generally applicable to environmental improvements and decrements and to all items that might be valued using nonmarket valuation methods.

The basic problem for reliability and validity assessment is that economists cannot see inside people's heads, which means that true WTP_is are unobservable. Choices revealed in markets or responses to survey questions provide proxy information about people's preferences that economists use to develop estimates of true values. It is this measurement process that gives rise to issues of validity and reliability, and the fact that WTP_i for each individual is unobservable makes assessing accuracy a complex task.

The inability to observe true values is a common problem in the social sciences. Measurement nearly always involves empirical steps designed to quantify theoretical constructs that cannot be observed directly. An IQ test is used to try to measure human intelligence, an abstract concept. Attitude questions in surveys are used to try to measure concepts such as racial prejudice and self-esteem, mental states that cannot be observed directly. Mathematics exams aim to measure competence in mathematics, an unobservable concept. Likewise, travel cost models, contingent valuation, and the other nonmarket valuation methods described in this

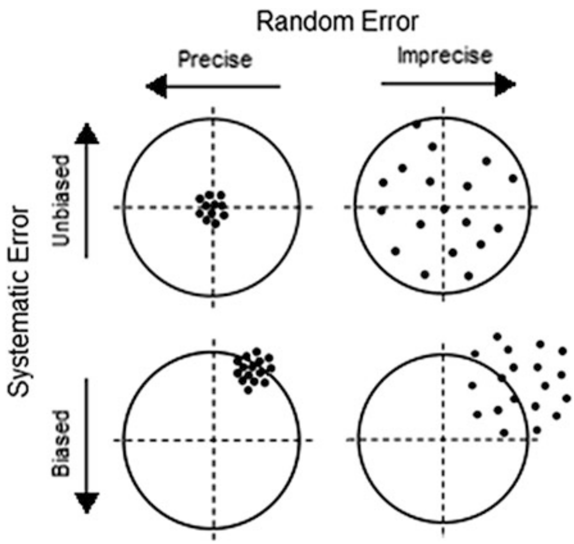


Fig. 12.1 Reliability and validity illustrated. *Source* https://www.ucl.ac.uk/ich/short-courses-events/about-stats-courses/stats-rm/Chapter_9_Content/relation_between_reliability_validity

volume are used to estimate the unobservable true values people hold for a change in environmental amenities or other goods or services.

If human intelligence, racial prejudice, WTP, and other such concepts could be observed directly, social scientists would simply go out and measure them. But since they are unobservable, researchers seek empirically tractable measurement procedures. Part of the research process is to investigate the accuracy of the measurements obtained using those procedures; this is “reliability and validity assessment.”

The above discussion focused on WTP at the individual level, but reliability and validity assessments are made at the sample level using summary statistics. To consider these concepts further, reliability and validity are illustrated in Fig. 12.1 using the metaphor of circles with dots. Suppose a value is estimated using one of the methods described in this book. A single WTP estimate can be thought of as one of the dots on a circle. The center of each circle represents the true value.

Now suppose that the research process is repeated and produces a new estimate of WTP. This new estimate is represented by a new dot on one of the circles. If this were done repeatedly, something like one of the four patterns in the figure would be obtained. If the pattern of dots forms a tight group somewhere on the target—as in the two images on the left side of the figure—the research procedures in question are providing estimates that are more reliable than those in the right hand images, which display larger dispersions of dots.

Going a step further, if the dots are scattered fairly evenly around the center (true value) as in the two circles at the top of the figure, the research procedures in question provide more valid (unbiased) estimates, while biased estimates are depicted by the two images on the bottom where the dots are not centered on the bull’s-eyes.

Figure 12.1, then, illustrates four possible situations. The research procedures in this stylized example could be reliable and valid (upper left), unreliable but valid (upper right), reliable but not valid (lower left), and unreliable and not valid (lower right). Obviously, ideal research procedures are both reliable and valid, but reliability and validity may not be as clearly demonstrated as presented in the figure. Nevertheless, some research procedures are more reliable (contain less variability) and more valid (engender less bias) than others. In the end, for different applications of any set of nonmarket valuation procedures, one may ask how much variance and how much bias are acceptable. In fact, the well-known variance-bias trade-off in statistics comes into play in the design and conduct of nonmarket valuation studies, where enhancing validity might increase variance or vice versa.

Like any metaphor, this one can become confusing if pushed too far. In particular, portraying reliability, and especially validity, in two dimensions as in the figure could lead to confusion. Only vertical distances are relevant when estimating WTP; dots above the horizontal line through bull’s-eyes represent positive errors in measurement and dots below represent negative errors, representing over and under estimates of WTP. Vertical lines with the true WTP in the center would be more correct but also seem less expressive.

Let us use an example to further explain. Suppose the research problem is to evaluate mean WTP for an improvement in water quality in a river. Let WTP^T represent a typical household's true WTP for the improvement.

Researchers collect data on property sales in the area that would be affected by the change in water quality and apply a hedonic property value model to estimate the marginal value for a typical household. Let this be WTP_j^E, which represents the estimated average value per household for a small improvement in water quality from the *jth* application of the hedonic procedures. In terms of shooting arrows at targets, WTP^T is the center of the target, and WTP_j^E is a black dot associated with the *j*th application. The relationship between the true value and the estimated values can be written as

$$\mathrm{WTP}_j^T = \mathrm{WTP}^E + e_j \quad \forall j, \tag{12.2}$$

where e_j is the error in estimating the true value and can be positive, negative, or zero. In fig. 12.1, the error is the distance from the center of the circle to the *j*th black dot. Accuracy, therefore, deals with the magnitude of e_j.

Estimation nearly always involves errors, so it is unrealistic to expect that $e_j = 0, \forall j$. Rather, the larger the expected error, the more likely it is that a valuation estimate may be deemed unreliable or invalid. Let us partition the error term as $e_j = e_{rj} + e_{mpj}$, where e_{rj} denotes the random component of the estimation error and e_{mpj} denotes measurement error associated with the *j*th application. The goal is to choose an estimation method (m) and procedures (p) such that e_{mpj} is as small as possible. The hedonic method selected might be a repeat-sales model to avoid issues of unobserved variables biasing value estimates, and the procedures would follow those outlined in Chap. 7 for estimating a hedonic model.

Focusing first on reliability, it is important to separate normal sampling variation from additional variation in estimates due to the valuation method and procedures used. Think of WTP_j^E as an estimated mean. Then the standard error of this estimate is

$$\mathrm{se}_{\mathrm{WTP}_j^E} = s_j \big/ \sqrt{n_j}, \tag{12.3}$$

where s_j is the sample standard deviation and n is the sample size. For purposes of exposition, let us partition the standard deviation into two sources of variation:

$$s_j = s_{rj} + s_{mpj}. \tag{12.4}$$

Here, s_{ij} is the variation due to variation across people (i) in the sample, and s_{mpj} is variation due to measurement error associated with the valuation method chosen and the steps selected to estimate WTP_j^E. For example, a hedonic data set that contained a larger number of outlier observations, both low and high sale prices, could increase dispersion of value estimates (let us assume, without introducing bias). Thus, an image such as portrayed in the upper right corner of Fig. 12.1 can occur because there is a lot of normal variation across sold properties (s_j). It is the case that $se_{\mathrm{WTP}_j^E}$ can be reduced by increasing sample size or by removing outlier observations from the data, thereby minimizing the effect of both sources of error and enhancing reliability. However, reliability can also be enhanced by following the study implementation steps discussed in Chap. 8 that avoid procedures that will increase dispersion in value estimates (s_{mpj}). Of course, all valuation methods allow for some measurement error, but to the extent that the steps used to measure values tend to avoid dispersion, the resulting value estimate is more likely to be deemed more reliable and therefore more accurate.

Turning to validity, this component of accuracy is associated with the expected value of e_j. If the $E\left(e_j\right) = 0$, which implies that measurement error does not lead to over- or underestimation, on average, then the valuation approach is unbiased and is valid. It is important to recognize that measurement error does not always imply bias. If the error is random with an expected value of zero $[E(e_{mpj}) = 0]$, the error only affects reliability. Validity concerns arise when the expected value of e_{mpj} is nonzero; bias is present in the estimation. The presence of bias does not imply an estimate is invalid because most estimates can contain some bias.

In the hedonic model example, bias could occur if an omitted relevant variable were correlated with some included variable used to estimate marginal WTP, which would bias the estimated coefficient. Similar to the bottom line for reliability, many valuation approaches involve some bias, and the larger the bias, the less likely it is that the valuation approach will be considered to provide valid estimates of WTP.

Now assume, for purposes of a thought experiment, that one might generate many samples of property sales using a jackknife approach where individual observations are sequentially removed from the data. The hedonic model is re-estimated for each jackknife data set, generating multiple estimates of WTP_j^E $(j = 1, 2, 3, \ldots)$ with associated errors. Each new application is another dot on the figure with an associated standard error.

Some errors, such as outlier observations that might be exposed with the jackknife robustness exercise, may only affect the variance of value estimates and have an expected value of zero (i.e., $E\left(e_j\right) = 0$). If this is true, then such errors merely add noise to estimates of WTP rather than bias, and thus the errors affect reliability but not validity. On the other hand, errors such as those associated with omitted relevant variables might not have zero expected errors $[E(e_j) \neq 0]$, and therefore could result in biased WTP estimates, which affects validity.

These are simple examples of factors in a hedonic study that might affect the reliability and validity of welfare estimates. At the core of this primer are steps analysts should follow when applying each of the valuation methods discussed. For the estimation of hedonic models, the example used in this discussion, the steps are outlined in Chap. 7.

Many of the steps for each method require decisions by analysts that can potentially affect the reliability and/or validity of the resulting value estimates. In doing valuation research, one does not normally have the opportunity to make a large number of measurements to assess the reliability and validity of results from each individual study. Researchers usually get only one shot at the target or at most a few shots to conduct robustness analyses of the accuracy of valuation estimates.

Robustness analyses can take many forms, and two types of robustness are considered here. In the first instance, there may be more than one recommended alternative to implementing a specific step associated with a valuation method. In the second, the implementation of a step can be left purely to investigator discretion. In both of these cases, robustness analysis provide transparency of the impact of investigator decisions on value estimates and can provide insights for future research to enhance the accuracy of valuation methods. Thus, research that is specifically designed to evaluate reliability and validity has a special role to play because it helps researchers conducting applied studies to judge how the choices they make in applying a selected method will affect the accuracy of the resulting value estimates.

The above discussion provides the bare bones of a framework for assessing the accuracy of nonmarket value estimates. The next section fleshes out the framework, a necessary step before illustrating how to apply it using the contingent valuation and travel cost methods.

12.2 Reliability and Validity Assessment

A valuation "method," as the term is used in this chapter, consists of a broadly defined set of procedures that can be used in estimating nonmarket values. Examples include the contingent valuation, travel cost and hedonic price methods. Most of the chapters in this book are devoted to explaining the basic procedures used when applying common valuation methods.

"Steps" are the various procedures that might be used to implement a method. Examples of steps include selection of a random-utility model to implement the travel cost method, the use of a referendum question to implement the contingent valuation method, or choosing a meta-analysis to implement the benefit-transfer method.

"Individual applications" take a method and apply a set of steps to estimate one or more nonmarket values.

12.2.1 Reliability and Validity of Methods and Individual Applications

Reliability and the three C's can be applied to individual applications, and are a function of the steps used to implement a study. Reliability and the three C's can also be used to consider whether entire methods are capable, if properly applied, of producing value estimates that are sufficiently accurate to be useful. Reliability and validity assessment of individual applications and entire methods are closely related. Much of what is known about the accuracy of methods comes from individual applications. Below, reliability and validity of individual studies are discussed first, and then the reliability and validity of methods as a function of the outcomes of many of such individual studies.

12.2.2 Reliability Assessment for Individual Applications

Reliability assessment for an individual application involves more than assessing whether an econometric estimator that minimizes variance was chosen; it includes assessing all steps in a study that can influence the magnitude of variance even when a variance-minimizing estimator is applied.

In the broader social science literature, reliability is typically investigated using a test-retest experiment where a set of steps is applied to implement the method at time t, and then those steps are replicated at time $t + 1$ (see Carmines and Zeller, 1979, and Zeller and Carmines, 1980). If the two estimates of value (t and $t + 1$) are statistically indistinguishable, the value estimation procedure is deemed reliable. By consequence, the chosen valuation method and implementation steps in a nonmarket valuation study are deemed reliable if they pass a test-retest criterion. In studies involving surveys (e.g., stated preference studies and most travel cost studies) the survey could be administered at two points in time. For methods that involve building data sets from sources other than surveys—most notably, many hedonic studies—data could be gathered at two or more points in time.

However, investigating reliability in nonmarket valuation studies can be tricky. For studies involving surveys, for example, the ideal design would be to have the the original sample of people retake the survey at $t + 1$. Hopefully, the slate would be wiped clean between the treatments, but this could be difficult to accomplish. The longer the duration between the test and retest, the less likely people will mimic their prior responses, but as the time between treatments increases, there may be fundamental changes that could affect preferences. For example, if time t was before the Great Recession and $t + 1$ was during, finding a significant difference in value estimates might not indicate unreliability.

While a more reliable valuation study is one that has a small variance associated with the estimated value (the left quadrants of Fig. 12.1), a small variance is not fundamental to test-retest experiments. That is, studies with small and large variances can be reliable using a test-retest investigation as long as the null hypothesis of no difference in value estimates between time t and $t + 1$ cannot be rejected. In fact, a large variance could make it more difficult to reject test-retest reliability, and a small variance could make it more difficult to confirm test-retest reliability.

In closing this section, the social science literature sometimes views reliability and test-retest reliability as synonyms. Because of the challenges outlined, it would be unfortunate if this usage carried over to nonmarket valuation research.

12.2.3 Validity Assessment for Individual Studies

Given that true values are unobservable, indirect evidence of validity (i.e., evidence that does not require that true values be known) must be used. Three types of validity are commonly considered: content, construct, and criterion validity (Carmines and Zeller 1979; Zeller and Carmines 1980)—referred to as the "three C's" of validity here.

As applied here, **content validity** is about procedures used to implement an individual study. Content validity assessment asks whether the valuation method chosen and all steps used to implement it are fully conducive to measuring the true value. This concept of content validity is broader than elsewhere in the nonmarket valuation literature. For example, Mitchell and Carson (1989) focused on the adequacy of the survey instrument; here the definition includes the survey, if any, but goes beyond it to include all the steps from the initial definition of value through the reporting of value estimates.

An example from outside nonmarket valuation will illustrate. Suppose researchers seek to measure the value of a market good using estimated supply and demand functions. When analysts or reviewers examine the procedures used in modeling supply and demand, they are considering the content validity of the application, as that term is used here. If the theoretical framework underlying the study is flawed, the data contain measurement error, the supply and demand equations are misspecified, econometric estimation procedures are poorly conceived, or other questionable procedures are identified, the ability of the study to adequately estimate the true value would be suspect, and content validity might not be established.

The same sorts of procedures are involved in assessing the content validity of nonmarket valuation studies. Various chapters in this book have highlighted the theoretical frameworks for value estimation and the data collection procedures and other steps to be taken in completing such studies. Reading what the authors said about how to implement these steps reveals a lot about what the literature implies regarding content validity. Researchers and analysts can use the recommended steps as the basis for designing studies that will enhance the content validity of the resulting value estimates. Those who read nonmarket valuation studies—whether as formal reviewers, scholars, or the users of the study results—can use the material presented in this book to assess the content validity of the study under consideration.

In making content validity assessments, it is important to recognize that there are often trade-offs that must be considered in study design and implementation. Thus, failure to follow any individual step recommended by the authors in this book—while reason to question content validity—is not sufficient to reject the content validity of the study. Investigators should document why the steps in study implementation were chosen or not chosen, and readers of the study should evaluate content validity globally within the documented context.

Moving from content to construct validity, the focus changes from procedures to the study's data and results. **Construct validity** assessment begins with prior expectations about how the true value ought to be related to other variables. Such prior expectations are motivated by theory, intuition, and past empirical evidence. They are translated into hypotheses that can be tested using the study's data and results.

Continuing the market demand and supply example, suppose the study estimates the demand for and supply of meat products in the U.S. Theory suggests there should be substitutes for beef, and through intuition and past evidence one would expect pork to be a substitute. Hence, one would expect the sign of the cross price effect between beef and pork to be positive. If estimation of the model reveals that this relationship holds, then the validity of study results would be supported. Not finding a positive sign would raise doubts about the study's validity.

Similar steps are followed in assessing the construct validity of nonmarket valuation studies. For example, holding other things constant, one would expect the price of a dwelling unit to be positively related to its square footage. Should a hedonic price function fail to show that relationship, this would be grounds for questioning the study's construct validity.

A common type of construct validity test involves "convergent validity." Convergent validity considers whether two or more valuation methods or approaches that are designed to estimate the same value provide statistically similar value estimates. Failure to find statistically similar value estimates would raise doubts about validity.

An example of a convergent validity test is comparison of value estimates from travel cost and contingent valuation studies (e.g., Hanley 1989), when both types of studies are designed to measure the same true value. If values from the two methods are statistically indistinguishable, this is evidence of construct validity. At the same time, it is not possible to conclusively say the value estimates are valid because both value estimates could contain biases in the same direction. Likewise, finding a statistically significant difference in value estimates is not conclusive evidence of invalidity.

The degree to which either study is invalid is not known because the true value is not known. It could be that one valuation study produced a biased estimate of value and the other did not, or that both studies provided biased estimates (e.g., one might have provided an overestimate and the other might have produced an underestimate). However, if the estimates are statistically indistinguishable, this is suggestive evidence but not confirmation that both methods have produced unbiased estimates. In sum, convergence is supportive of validity but not conclusive.

Cross validation is and another type of construct validity, which involves using an estimated model to predict observations for a hold-out sample. This type of validity assessment, for example, would be used in robustness analyses to evaluate the econometric models used to analyze nonmarket data.

Criterion validity tests involve comparing results from two valuation methods: one method with uncertain validity and another method that is widely accepted as having a high level of validity. Results from the widely accepted method serve as the "criterion" (the proxy for the true value) to evaluate the accuracy of results from the method with uncertain validity. Of course, if the method with widely accepted validity could be broadly applied at a reasonable cost, it would be used and there would be no need for the method with uncertain validity. However, the more accepted method might be expensive or only usable under special circumstances.

Consider an example from the stated preference literature. List and Gallet (2001) summarized criterion validity tests for what has come to be known as "hypothetical bias" in stated preference studies. In laboratory or field settings, results from stated preference studies were compared with results from transactions involving the actual exchange of money for goods or services. Since most economists would be satisfied if values from stated preference studies were comparable to values from well-functioning markets, results from the cash transactions serve as the criterion for judging the validity of results from the parallel stated preference applications where cash and goods or services are not exchanged.

If stated preference studies perform well in comparison with cash transactions, this is evidence of criterion validity. The opposite also holds. If stated preference studies perform poorly in comparison with cash transactions, this is evidence of invalidity. It is only possible to say this is evidence of validity or invalidity because the presumption that the criterion is an unbiased measure of the true value might or might not be correct, and also because criterion validity—like other accuracy measures—is really a matter of degree rather than an either-or test. The discussion will return to hypothetical bias in the section on the validity of contingent valuation.

Confusion occasionally arises about whether some tests are construct validity or criterion validity tests. For example, comparing the results from contingent valuation studies with results from travel cost studies investigates construct validity. Calling comparisons of contingent valuation and travel cost estimates "construct validity tests" implicitly assumes that neither is considered more valid than the other. In contrast, in the criterion validity example, the outcomes from cash transactions have presumed validity, while stated preference methods have uncertain validity. Thus, while neither construct nor criterion validity tests are perfect, criterion validity tests are stronger indicators of validity because the criterion measure is presumed to be unbiased (or at least contain minimal bias).

12.2.4 Reliability, the Three C's, Judgment, and the Weight of Evidence

All empirical studies contain random variability in estimation and some bias is often unavoidable. There are no litmus criteria to say how much variability makes a study unreliable and how much bias is required to make a study invalid. That is, variance and bias are statistical concepts, whereas reliability and validity are judgments that are imposed on statistical observations. Such judgments are where disagreements arise because various people can apply differing standards with regard to what is acceptable.

When valuation studies are conducted to support decision-making, the debate about accuracy centers on how much bias and unreliability are tolerable for the value estimates to be of practical use. This type of debate is good as long as it centers on the observable facts and not on goals of justifying a personal or political position. Debates about reliability and validity should take place within the concepts discussed above. For example, reaching judgments about reliability should take into consideration the amount of variation that is acceptable for the decision the estimated values are intended to support. In benefit-cost analysis, it would matter if the variability in value estimates was such that—even under the most conservative value approach—benefits exceeded costs; this would be evidence of project feasibility. If this were not the case, there would be concern that the project might not be economically justifiable.

Reaching judgments regarding the validity of individual applications should involve consideration of the full weight of evidence across the three C's. The strongest validity inferences will apply to studies that (1) use methods and follow implementation steps with established content validity, (2) have passed appropriate construct validity tests, and, hopefully, (3) have chosen a method and followed steps that have passed criterion validity testing. The three C's can be thought of as the three legs of a stool. Judgments about the validity of value estimates should rest on all three legs. If one or more legs are weak, validity suffers but is not necessarily fatally compromised. A study with strong content validity might be judged less valid if prior expectations are not met in convergent or other construct validity tests. Likewise, an individual application might employ methods and steps that have performed well in criterion validity testing, yet it could be judged less valid if content and/or construct validity criteria are not met. For example, a state-of-the-art, unbiased econometric estimator might be applied to the data, but if the data collection step created flaws in the data, the resulting value estimate could still be biased and deemed invalid.

The bottom line is that reliability and validity are not clear-cut, yes-or-no issues. They require careful attention to the weight of all the evidence.

12.2.5 Drawing Conclusions About the Validity of Entire Methods

The functioning of the three C's as mutually reinforcing legs of a stool can clearly be seen in the process through which new methods either achieve success or are abandoned. Early construct and/or criterion validity testing of new methods yields insights about what procedures succeed and what do not. In turn, these early insights are implemented and tested in more studies. As the process continues, the three legs upon which a successful method rests either become progressively stronger or the method is abandoned.

A method is more or less content valid based on whether a body of researchers agrees on a set of steps that can be used to estimate true values to a satisfactory approximation. The set of steps constitutes the state of the art in applying the method. Judgments about the accuracy of a method are based on accumulated evidence. Consensus among researchers about the content validity of a method is based to a large degree on construct validity studies published in the peer-reviewed literature.

Turning to construct validity, a method can be judged to be more or less construct valid based on the extent to which an accumulated body of results from individual studies is fairly consistent in meeting prior expectations based on theory, intuition, and results from earlier research. The caveat "fairly" means that some failures of construct validity tests in individual applications are not damning. Such failures can result from peculiarities in the circumstance of the errant study, statistical flukes, going beyond the domain where the method can be applied, or other causes. Or a failure might hold the key to improving the steps in planning and implementing the method. On the other hand, if the method in question fails such tests frequently, this would point toward the conclusion that the method is not construct valid. Either the method will have to be improved or abandoned.

Typically, construct validity tests focus on a particular step in the valuation process in a specific situation or context. Replications will normally be needed to see if test results are robust in similar contexts and across contexts. If the method in question is based on stated choices, for example, and it seems to test out well in a water quality study, is this success robust across other water quality applications? Does it work well in air quality studies? Does it work well in both developed and developing world applications?

If criterion validity tests prove feasible, all the better. Criterion validity tests can, in principle, be particularly potent in evaluating the validity of a method. But again, failure in a few studies might not be fatal. Furthermore, replications are needed to see if the specific tests are robust. There is also a danger that both the criterion study and the method in question are flawed and that the flaws offset each other, yielding a false positive result. Replications of criterion validity tests should help identify such problems.

These conditions have much in common with the rules for admission of expert testimony on a wide variety of topics (including nonmarket values) into U.S. court proceedings. The courts apply what is called the Daubert standard from a case in federal court, Daubert v. Merrell Dow Pharmaceuticals Inc., 509 U.S. 579 (1993), as well as later federal cases. In considering whether to admit expert testimony, courts are to weigh whether the technique or theory applied has been subject to peer review and publication and whether the technique or theory has been generally accepted in the scientific community (Blinka 2011).

Saying that a body of researchers achieve some level of consensus does not mean that there will be complete agreement. Within the body of researchers who have engaged in the development of a method, debates will likely continue about how to apply the method correctly and how to improve the method. There might also be a body of critics that continues to believe that the method lacks validity. The "contingent valuation debate" will be considered as a case of this kind in the next section. This is a healthy part of the evolving science as long as everyone involved agrees in principle to base their judgments on the weight of all the evidence and not on preformed opinions. In the end, what matters is the judgment of the weight of evidence.

From the foundation that has been outlined, the rest of this chapter is devoted to considering how reliability and validity might be applied in evaluating two well-known valuation methods. To illustrate applications to stated preference methods, contingent valuation is used as an example. Then travel cost is used as a revealed preference example. The chapters on these methods (Chaps. 4 and 6, respectively) present the steps involved in doing state-of-the-art studies. The following sections consider whether contingent valuation and travel cost studies, using state-of-the-art steps, are capable of yielding reliable and valid estimates of WTP. While the ultimate goal is to judge whether the value estimates coming out of individual studies are sufficiently reliable and valid, focusing on methods is a necessary preliminary activity. One could not hope to attain accurate value estimates applying a method that is known to be unreliable or invalid.

12.3 Accuracy of the Contingent Valuation Method

We are using contingent valuation as a case example for stated preference methods, but the accuracy framework has general applicability to other stated preference methods. For example, valuation question formatting is relevant to all stated preference methods, but the specifics of question formatting will vary by method.

12.3.1 The Contingent Valuation Debate

Before proceeding to the reliability and validity of contingent valuation, it is necessary to acknowledge the substantial debate that has continued over the past 25 years about the credibility of contingent valuation estimates—and stated preference value estimates more generally, for that matter—in legal, policy, and academic settings. The trustworthiness of contingent valuation has been seriously challenged since the settlement of monetary damage claims against Exxon in the aftermath of the Exxon Valdez oil spill. Based on growing evidence that contingent valuation produces valid estimates of value (see Mitchell and Carson, 1989, for a summary of the evidence at that time), both the federal government and the state of Alaska initiated contingent valuation studies to estimate monetary damages that occurred due to the oil spill. After settlement of the case (without explicit reference to any contingent valuation study), a number of journal articles from members of Exxon's team of experts, including Diamond and Hausman (1994) and McFadden (1994), as well as an edited volume with various chapters from the Exxon team (Hausman 1993), challenged the validity of contingent valuation to estimate values to support legal claims and decision-making. Their conclusion is succinctly summarized by the statement, "We conclude that the contingent valuation method does not measure an economic value that conforms with economic preference concepts" (Diamond and Hausman 1993, p. 4).

The critique has many dimensions, and the debate continues today. Going into detail here would be a detour from the goals of this chapter, but some examples will illustrate. The critics emphasize what came to be called "embedding effects," which refer to the findings in several contingent valuation studies that an item has different values depending on whether it is valued independently or as part of a package (see Kahneman and Knetsch 1992). In addition, results from verbal protocol studies, where subjects were asked to speak aloud about what they were thinking as they answered the valuation questions, were interpreted as antithetical to the revelation of economic preferences (Schkade and Payne 1994). It was also argued that contingent valuation responses include "warm glow," and that warm glow should not be counted as an economic value (Diamond and Hausman 1994). A more recent branch of the debate centers on whether value estimates exhibit sufficient responsiveness to the resource changes valued in scope tests (Desvousges et al. 2012).

These critiques have not been conducted using a holistic accuracy framework as is proposed here, and they improperly rely on results from one or a few studies to attack the generally method without considering the full weight of the evidence. To set a broader context here, the discussion considers what can be gleaned, on a more general level, from the literature regarding the reliability and validity of contingent valuation. A big picture perspective based on the weight of evidence is needed.

12.3.2 Reliability of Contingent Valuation

A cohesive literature has addressed the reliability of contingent valuation estimates using test-retest procedures. Early studies (e.g., Jones-Lee et al. 1985) resurveyed subsamples after some period of time had passed. Responses at the individual level were compared for consistency. Going back at least to Reiling et al. (1990), investigators realized that resurveying the same subjects could be misleading if subjects remembered their first responses when completing the second survey. However, McConnell et al. (1998) addressed this question econometrically and concluded that correlations between earlier and later responses were due to heterogeneous preferences, rather than to recalling earlier responses.

A large number of studies, including those cited in the preceding paragraph, found evidence to support the conclusion that well-conducted contingent valuation studies produce reasonably reliable estimates of value (see also, Berrens et al., 2000, Carson et al., 1997, Kealy et al., 1988, 1990, Loehman and De, 1982, Loomis, 1989, 1990, Reiling et al., 1989, and Teisl et al., 1995).

12.3.3 Validity of Contingent Valuation

Is contingent valuation a valid valuation tool? This is a question that has been asked repeatedly, most notably in the contingent valuation debate. Skepticism about stated preference methods is ingrained in the culture of economists. Most economists appear to have an unfailing faith in the validity of revealed preference methods and tend to distrust surveys as sources of data on preferences. However, looking deeper reveals that results from contingent valuation studies might be more accurate than some would like to portray.

12.3.3.1 Contingent Valuation Content Validity

The steps laid out in Chap. 4 represent procedures, based on the evidence in the literature, that most researchers in the field would find satisfactory and would tend follow in the design and implementation of a state-of-the-art contingent valuation study. This is not to say that consensus is complete. Far from it. To quote the proverb, the “devil is in the details.” But few of the researchers cited in Chap. 4 would disagree with the general efficacy of carefully defining the changes to be valued, deciding whose values are to be counted, defining a suitable sample size, and so on. And there seems to be substantial agreement on many of the details under each step. For example, few would question the desirability of subjecting survey drafts to qualitative research and pretests, framing questions as referenda where possible, collaborating with scientists outside of economics, designing scenarios to be incentive-compatible, and applying appropriate econometric procedures. To the extent that disagreements about exact procedures persist, they are grist for the research mill.

On the other hand, while following the steps for implementing a contingent valuation study outlined in Chap. 4 can enhance the content validity of a study, there are choices that investigators must make routinely where there is no clear-cut guidance from theory or the literature, and these choices can influence value estimates.

Three examples come immediately to mind. First, the issue of what payment vehicle to use persists. Many recent studies have used tax vehicles, but this remains an uneasy compromise. Given the aversion many people have to increased taxes, the effect of using this type of payment vehicle is likely to be a downward bias (payment vehicle bias) in value estimates, all else being equal. And there are situations where taxes are implausible or otherwise unworkable. Then what?

Second, the appropriate frequency of payments and their duration is not at all clear. Theoretically it should not matter, but anyone who has conducted a focus group for a contingent valuation survey will tell you that it matters whether payments are monthly or annual and whether they one time or extend over some other period of time. This is not simply a matter of discounting at some conventional rate (Stumborg et al. 2001).

The third example is the issue of whether to estimate individual or household values. While conventional theory focuses on the individual consumer, it is obvious that many households engage in joint budgeting and decision-making. There is no clear guidance currently from theory or from empirical studies about how to resolve this dilemma. In applied studies, investigators often choose to make relatively conservative assumptions, that is, assumptions that tend to yield relatively low values. This boils down to introducing assumptions that, if anything, bias values downward. While practically expedient, intentionally introducing negative biases compromises content validity and increases the likelihood of rejecting some utility-enhancing projects or actions. Until such issues are resolved, the content validity of contingent valuation applications is enhanced by clearly documenting investigator decisions with support from the published literature or information gleaned from focus groups, and by conducting robustness analyses where possible to understand the potential magnitudes of the impacts on value estimates.

In sum, the literature supports a case for concluding that the contingent valuation method has substantial content validity while admitting that research needs to continue to strengthen the method and that, in the meantime, documentation of procedures and checks for robustness are needed.

12.3.3.2 Contingent Valuation Construct Validity

Economic theory in its most general form suggests two hypotheses that can be tested when the contingent valuation method is applied. The first might be loosely termed "negative price sensitivity." Stated differently, theory implies that the demand functions, implicit in contingent valuation responses, be downward sloping. The second might be termed "positive income sensitivity." A logical question to ask is whether estimates are positively related to respondents' incomes (i.e., do people with higher incomes express higher values on average?).

Price sensitivity can be illustrated using responses from dichotomous-choice questions. As the offer amount increases, the probability of an affirmative answer should decline. This expectation is less easy to grasp when contingent valuation questions ask respondents to directly reveal their full WTP values, such as with open-ended questions. Here, negative price sensitivity follows when different respondents express different WTP amounts. Suppose there are only two respondents who state two values—WTP_1 and WTP_2—in response to an open-ended question where $WTP_1 > WTP_2$. One can infer a demand function for the public good in the following way. If some price, p_1, such that $WTP_2 > p_1$, could be charged, both respondents would demand the good. At a higher price, p_2, such that $WTP_1 > p_2 > WTP_2$, only one would demand it. If the price were even higher at p_3, where $p_3 > WTP_1$, neither would demand it. In this way, negative price sensitivity could be established. Few if any contingent valuation studies have failed price sensitivity tests.

The construct validity of contingent valuation is further supported by the fact that many studies have found that income has a positive effect on values. For example, Jacobsen and Hanley (2009), in a meta-analysis of WTP for biodiversity, found that values rise with income and that the income elasticity of WTP is less than one (see also Schläpfer, 2006). While one desires income responsiveness in individual studies, one would also expect income elasticity to vary across studies, which both Jacobsen and Hanley, and Schläpfer found.

At the same time, some individual applications have failed to find positive income effects. What are we to make of this? Failure to find a positive effect of income in an individual application does raise some concerns about the validity of that individual study. However, how concerned one should be about lack of an income effect in individual studies is a matter of judgment. On the one hand, expecting a positive income effect has a strong foundation in theory. On the other hand, four issues need to be considered. First, the good being valued could be an inferior good or simply neutral with respect to income. Second, income is difficult to measure well (Moore et al. 2000). Third, accurate measures of respondents' incomes might not be what are needed. After all, in a broader sense, WTP is likely to be based on ability to pay, which can involve wealth as well as income. And fourth, in countries like the U.S., most values from contingent valuation studies involve less than 1% of median household annual income. Decision-making to allocate and reallocate such small amounts of discretionary income is not well understood by economists.

At any rate, failure to find a positive relationship between income and value estimates in some studies is not grounds for questioning the construct validity of the overall contingent valuation method. If lack of income effects were inherent in the method, then most or all studies would fail to find positive income effects. This is not the case.

Scope tests are another type of construct validity test that can be applied in contingent valuation studies. A scope test asks whether value estimates are sensitive to changes in the magnitude of the amenity, good, or service being valued. Again, there is a close link to economic theory. More of a desired item should have a higher value. The possibility that scope test failure is inherent in contingent valuation was first raised with Kahneman and Knetsch (1992) and gained momentum with Desvousges et al. (1992), which was part of the Exxon studies. The latter study found that value estimates were the same for preventing 2,000, 20,000, and 200,000 bird deaths.

As with income effects, scope test failures raise questions about the construct validity of individual studies, but again, there can be extenuating circumstances. For example, Heberlein et al. (2005) found that respondents from northern Wisconsin were willing to pay less, on average, to increase the wolf population of the region to 800 animals than to increase it to 300 animals. Postsurvey interviews showed that many respondents were in favor of the smaller increase, but feared that the larger number might have too many ill effects on wildlife, domestic animals, pets, and even people.

Furthermore, statistical identification of a scope effect can be empirically challenging. For example, between-subject scope tests are viewed as being the most credible type of investigation. In such tests, one sample sees one change in a good or service, and a separate, independent sample sees a larger or smaller change. Then, the value estimates from the two samples are compared. This forces a contingent valuation study to meet a standard that even market valuation studies are not subject to. In any market setting, consumers view different quantities and qualities with different prices rather than only a single price and quantity, as in a split-sample scope test. Despite such a high bar for a credible scope test, the weight of evidence suggests that most contingent valuation studies do pass a scope test (Carson 1997; Desvousges et al. 2012). Hence, the evidence does not support the hypothesis that scope test failures are inherent in the method.

Price, income, and scope tests are weak tests. Such effects could have the right sign and yet be incorrect about the magnitude. For example, the fact that a study passes a scope test does not mean that the magnitude of value estimates or the estimated difference in the values is accurate. There have been attempts to devise stronger construct validity tests based on theory. Diamond (1996) proposed to test for consistency of estimated values under assumptions about the curvature of the utility function. Diamond and Hausman (1994) proposed to test for the adequacy of scope differences based on a so-called "adding-up" test (see also Desvousges et al., 2012, and Hausman, 2012). The justification for these tests remains controversial (see, for example, Haab et al., 2013), and a thorough treatment would take us too far afield. It must be sufficient to say that basic economic theory does not provide guidance for constructing such tests. The investigators must apply assumptions that impose structures on preferences—assumptions that are not based in the general axioms of revealed preferences and might not be justified. Passage of such tests enhances conclusions regarding validity, but test failure may or may not imply invalidity.

Next, recall that construct validity tests can be derived from intuition as well as from theory. Looking beyond price, income, and scope sensitivity, researchers have also found significant, intuitively plausible relationships between contingent values and respondent environmentalism, the use of the resource, the distance respondents live from the resource, and other such variables. Such variables can often be interpreted as proxies either for preferences or for differences in individual circumstances. Widespread demonstrations of expected relationships between WTP and these proxy variables have encouraged the conclusion that contingent valuation studies tap into economic preferences and, hence, that the method is adequately construct valid (Carson et al. 2001).

A number of studies have compared values estimated using revealed preference methods with contingent valuation where the studies were designed to measure the same or very similar values (see Carson et al., 1996). Hence, these studies provide evidence relating to convergent validity. The Carson et al. study is a meta-analysis of studies that provide both contingent valuation and revealed preference estimates of value (based on travel cost, hedonic, and averting methods). They found a total of 83 studies that supported 616 comparisons between contingent valuation estimates and revealed preference estimates for the same goods (many studies had several value estimates). They presented several statistical tests that tended to show that contingent valuation estimates are less than comparable revealed preference estimates. For example, when all the value comparisons were included, the ratio of contingent valuation estimates to revealed preference estimates averaged 0.89 (95% confidence interval = 0.81–0.96). With a data set trimmed for outliers, this ratio fell to 0.77, and the null hypothesis that the ratio was equal to unity was soundly rejected.

Two other meta-analyses have not limited themselves to studies that contain both contingent valuation and revealed preference estimates of value but have instead asked simply whether, across a range of studies, contingent valuation tends to yield higher or lower values than revealed preference studies for similar goods or services. Kochi et al. (2006) found that contingent valuation-based values of a statistical life average substantially less than values from hedonic wage studies. Rosenberger and Loomis (2000) conducted a meta-analysis of outdoor recreation studies and also found that contingent valuation studies have a statically significant tendency to yield lower values than travel cost studies.

This tendency for contingent valuation to yield somewhat lower values than revealed preference analyses means that one approach or the other or both could tend to produce biased estimates. However, concluding that a bias exists might be premature. Remember that contingent valuation estimates are best interpreted as Hicksian values, while revealed preference values should be considered Marshallian values. Hence, if anything, revealed preference values of WTP from each of these methods may be different due to the different theoretical constructs measured.

Setting aside the Hicksian–Marshallian distinction, this may be an instance where the proverbial glass may be half (or more) full, rather than half empty. Yes, it would be cleaner if the two approaches yielded the same values. However, given that the approaches are so different, the fact that value estimates come out this close seems to be good news about the construct validity of the two types of methods. Carson et al.'s (1996) finding that revealed preference measures and corresponding contingent valuation measures are highly correlated further reinforces this conclusion. This suggests that revealed preference and contingent valuation methods are measuring the same underlying values, though some bias could still be present.

Therefore, looking across the various kinds of evidence suggests that the weight of evidence is tipping the scales in favor of the construct validity of the contingent valuation method.

12.3.3.3 Contingent Valuation Criterion Validity

Two types of criterion validity tests have been conducted: those using simulated markets and those using actual voting behavior. In simulated market experiments, values based on actual cash transactions serve as the validity criterion for comparison with contingent valuation estimates of value. In the voting comparisons, a contingent valuation study using a referendum format is conducted at approximately the same time as an actual referendum; here voting patterns in the actual referendum serve as the criterion. Both approaches can provide important insights.

Three meta-analyses have compared contingent valuation results to results from parallel simulated markets (List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005). All of these studies concluded that contingent valuation estimates tend to exceed simulated market estimates. For example, Murphy et al. used a data set with 28 studies that met three criteria: (1) only studies estimating WTP (not WTA) were used, (2) the same valuation mechanism had to be used in both the contingent valuation and simulated market treatments, and (3) the values compared had to be expressed in currency (not, say, in the percentage of people voting "yes"). Their analysis used the ratio of the contingent valuation estimates to the simulated market estimates, which is sometimes referred to as the "calibration factor." The calibration factors in the data set ranged from 0.76 to 25.08, with a mean of 2.60, a median of 1.35, and a standard deviation of 3.52. Assuming that the simulated market values are suitable criteria, the contingent valuation studies in these experiments overestimated WTP, a phenomenon that is generally categorized under the catchall title of "hypothetical bias."

But let us step back for a moment and ask how strong the case for hypothetical bias really is. The first thing to note is that the results of the hypothetical-bias experiments run counter to those relating to convergent validity cited in the preceding section. Comparisons of results from contingent valuation and revealed preference methods have rather consistently shown that contingent valuation estimates tend to be smaller, not larger.

Furthermore, the contingent valuation treatments in the hypothetical-bias experiments typically stress that the valuation exercises are completely hypothetical. Using the concepts developed in Chap. 4, the contingent valuation treatments in these studies are neither consequential nor incentive-compatible. Stressing that nothing will happen as a result of the subjects' responses makes the treatments inconsequential. Stating that the subjects will not under any circumstances have to actually pay makes the exercises incentive-incompatible.

A growing body of evidence supports the conclusion that hypothetical bias does not occur in well-designed contingent valuation studies that are consequential and incentive-compatible. Because many researchers prefer referenda, the branch of the hypothetical-bias literature dealing with referenda is particularly relevant. When referenda where the probability of actually having to pay is zero are contrasted with referenda where paying is certain and the item being valued will actually be supplied, hypothetical bias consistently identified (Cummings et al. 1997; Bjornstad et al. 1997; Taylor 1998; Cummings and Taylor 1998; Taylor et al. 2001; Brown et al. 2003; Landry and List 2007; Burton et al. 2007; Carson et al. 2014). Three of these studies also included treatments where the probability of actually paying was varied. Cummings and Taylor found that hypothetical bias declined as the probability of actually paying increased and disappeared altogether when the probability of paying reached 75%. Landry and List found that hypothetical bias disappeared when the likelihood of the referendum being binding was increased from 0 to 50%. Likewise, Carson et al. found that hypothetical bias disappeared when the likelihood of the referendum being binding was 20% and larger.

Of course, applied contingent valuation studies cannot provide probabilities of supply and paying, but most stress that policymakers will consider study results in decision-making. Most studies also diligently seek believable payment vehicles that could be implemented if the change being valued is provided to consumers.

Finally, the results of the hypothetical-bias literature are also in marked contrast to studies comparing actual voting with contingent valuation results. The studies that have conducted this type of comparison (Mitchell and Carson 1989; Champ and Brown 1997; Vossler and Kerkvliet 2003; Vossler et al. 2003; Johnston 2006) have shown that contingent valuation studies predict actual voting behavior rather well. Actual referenda are both incentive-compatible and consequential, and this seems to carry over to contingent valuation applications to public goods.

The conclusion, then, is that the hypothetical-bias literature is much less compelling on close examination than it was at first glance. It is important in underscoring the need to make contingent valuation survey scenarios incentive-compatible and consequential.

One word of caution should be inserted before moving on. The fact that hypothetical and real-value estimates are not statistically different does not indicate that this occurs because of any particular step in the valuation process. A step can enhance validity, reduce validity, or be neutral. Criterion validity assessment only provides insight of validity or invalidity for the entire package of steps used to develop the value estimate. Likewise, if invalidity is inferred, then carefully controlled experiments are needed on individual steps to find the source of the invalidity, ultimately returning to a new test of criterion validity. The evolving consequentiality literature is one example of this iterative process.

12.3.4 Contingent Valuation Accuracy and the Weight of Evidence

What can be said by way of conclusions about the reliability and validity of contingent valuation? The reliability of value estimates appears to be supported over a large number of studies. On the validity side, consider each of the three legs of the validity stool.

The **content validity** of contingent valuation is supported by extensive application of the method over many decades and across many countries, leading to considerable consensus about how to design and execute successful studies. Most notable in recent years is the increasing use of design features, such as dichotomous-choice questions, that are incentive-compatible and consequential. On the other hand, the heavy dependence on researcher choices suggests that much still needs to be done to investigate issues related to content validity. In the meantime, careful documentation of choices is necessary along with robustness checks.

There is considerable support in the literature for the **construct validity** of contingent valuation. Prior expectations involving price, income, and scope sensitivity are met sufficiently often to conclude that insensitivity to these variables is not an inherent flaw in the method. Sensitivity of values to environmentalism, environmental attitudes, past use of the resource, and the like provide further support. At the same time, the weaknesses of available construct validity tests must be acknowledged; having the expected sign does not reveal the presence or absence of bias in value estimates. Convergent validity tests, taken together, add further support for a conclusion that construct validity holds for contingent valuation, but her again additional research can enhance insights.

In terms of **criterion validity**, a growing number of studies provide evidence that hypothetical bias can be traced to elements of contingent valuation scenarios that are being addressed through incentive-compatible designs that are consequential. More studies like these to enhance design should be a high priority.

Consider again the contingent valuation debate. As a result of the controversy surrounding the credibility of contingent valuation, the U.S. National Oceanic and Atmospheric Administration convened the Blue Ribbon Panel on Contingent Valuation (often referred to simply as the NOAA panel), consisting of four prominent economists, including two Nobel laureates, as well as a widely respected survey researcher. While the panel acknowledged that the critics had raised important concerns, they ultimately concluded the following:

> Contingent valuation studies convey useful information. We think it is fair to describe such information as reliable by the standards that seem to be implicit in similar contexts, like market analysis of new and innovative products and the assessment of other damages normally allowed in court proceedings. Thus, the Panel concludes that contingent valuation studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use [nonuse] values. (Arrow et al. 1993, p. 4610).

The panel uses the term "reliable" in a much more general sense than it is used in this chapter. It is synonymous with the term "accuracy" here, which includes both reliability more narrowly defined and validity. The panel's use of "reliable" appears to follow the legal system's use of this term.

The NOAA panel made several recommendations that they consider essential for the accuracy of contingent valuation studies. For example, these studies should format their valuation questions as referenda, conduct scope tests, and include reminders of subjects' budget constraints. Nearly all of its recommendations fit in nicely with the framework espoused in this chapter and are incorporated in the best studies today.

Much research has been conducted since the NOAA panel concluded its review. Most of the results add further support to the case for concluding that well-conducted contingent valuation studies have sufficient accuracy (reliability and validity) to be useful for policy analysis and litigation.

12.4 Accuracy of the Travel Cost Method

This section applies the accuracy framework to the travel cost method. Using the accuracy framework, this section considers if the travel cost method, when applied using the state-of the-art recommendations in Chap. 6, is capable of producing reliable and valid estimates of true values.

The travel cost method is used as a case example of revealed preference methods, including the hedonic and averting behavior methods, where the accuracy framework will have general applicability. For example, considering the effect of model specification is relevant to all three revealed preference methods discussed in this text, but the specifics of model specification will vary by method.

Travel cost method practitioners have been concerned with accuracy issues for a long time. Such investigations can be traced back to some of the earliest travel cost studies (e.g., Brown and Nawas 1973; Cesario and Knetsch 1970; Stevens 1969; Trice and Wood 1958). However, past research has not been well organized around the fundamental concepts of reliability and validity. In fact, it appears that this statement applies to all revealed preference methods of valuation.

What is said here should be considered a first tentative attempt to evaluate what the empirical literature implies regarding the reliability and validity of the travel cost method, and by extension other revealed preference methods paralllel to the stated preference literature. Others will no doubt have much more to say.

Unlike contingent valuation, the travel cost method has been relatively free of controversy. In fact, the Exxon team published a travel cost study as a credible valuation exercise in the same volume where they adamantly criticized contingent valuation (Hausman 1993, Chapter VIII). But, as the following sections illustrate, there are many issues in applying the travel cost method. The lack of controversy surrounding the use of value estimates from travel cost models, is likely due to the acceptance of revealed preference data prima facie, with perhaps a blind eye to the many necessary investigator choices involved in the application of the travel cost method.

12.4.1 Reliability of the Travel Cost Method

Is the travel cost method a reliable valuation tool? This is a question that has not been asked directly. Some insights can nevertheless be gleaned from a handful of studies on recall of recreation behavior that provides data for the estimation of travel cost models. Several studies have compared data on reported recreational participation from independent samples varying the length of time for recall. The U. S. Fish and Wildlife Service conducts a periodic National Survey of Fishing, Hunting, and Wildlife-Associated Recreation that collects data that could be used to estimate recreation demand models. The traditional mode of collecting data for the National Survey had been to ask participants to recall their recreational visits over the preceding year. Chu et al. (1992) found that annual recall required in the National Survey led to over-reporting and increased variance of recreation participation relative to semiannual and quarterly recall periods. Gems et al. (1982) found that longer recall periods resulted in this same pattern of results for marine sport anglers who took four or more trips per year when two-week and two-month recall periods were compared, but variance did not increase for nonavid anglers (three or fewer trips). These findings were replicated in a study of recreational fishing where the standard deviation of days fished more than doubled for three-month and six-month recall periods versus for immediate recall (Tarrant et al. 1993). In these types of recall studies, it is presumed that data reported for shorter recall periods are more accurate. The collective findings from these studies suggest that using recreation-participation data with long periods of recall could tend to increase the variance of reported participation and hence reduce the reliability of the travel cost method, all else being equal. Over-reporting can also lead to "recall bias," which can affect validity, as described further on.

However, recall might not be a problem in all study settings. Mazurkiewicz et al. (1996) conducted a test-retest experiment using immediate and four-month recall for a one-week moose hunt, and found that the four-month recall period did not affect the mean or the variance of hunter participation. This finding does not necessarily contradict the finding of Chu et al. (1992), Gems et al. (1982), and Tarrant et al. (1993) because the moose hunt is a one-week hunt; thus, it is unlike fishing where there can be many shorter trips over several months.

The studies just discussed did not actually explore the implications of long recall versus short recall on value estimates from travel cost studies. With random-utility models, the focus is on choices among substitutes and the recall studies are silent on this important component of the estimation. Intuitively speaking, the travel cost method may have less reliability in applications with long recall periods than shorter periods.

One way to seek additional insights is to consider meta-analyses of travel cost value estimates. At least one study suggests that outcomes are systematically related to several resource and subject characteristics that one would expect to influence value estimates (Smith and Kaoru 1990a, b). If there were a high degree of variability in value estimates due to random error, it is unlikely that a statistical synthesis of the literature would reveal the strong statistical relationships found by Smith and Kaoru (1990a).

Still, recall studies and meta-analyses do not answer the fundamental reliability question of whether travel cost studies conducted with multiple samples from the same population would produce approximately the same value estimates—a tight cluster of dots versus a broad spread of dots as shown in Fig. 12.1. Further research is warranted on this topic.

12.4.2 Validity of the Travel Cost Method

Is the travel cost method a valid valuation tool? This is a question that has rarely been directly investigated because of the presumed validity of revealed preference data. However, looking deeper reveals a number of issues that remain at least partially unresolved.

12.4.2.1 Travel Cost Content Validity

Is there broad agreement among a body of researchers on the steps that should be followed in travel cost studies to achieve at least minimal accuracy in estimating true values? Evidently, the answer is "yes." The travel cost method is far from being considered an experimental method that has yet to prove itself. Rather, it stands as a high-ranking tool among revealed preference methods. Many members of the body of researchers are cited in Chap. 6, along with the large volume of peer-reviewed literature on the topic.

Still, a number of soft spots in current practice were identified in the conclusions of Chap. 6. For example, more realistic ways of valuing travel time are needed. Overnight trip and multiple-purpose trip modeling need to be improved. Practical, realistic models of intertemporal substitution need to be developed to account for the possibility that, if one cannot visit a preferred site today, one may, instead of going to a substitute site, visit the preferred site at a later date. The Kuhn–Tucker approach still needs to evolve and be more widely applied. Models need to be expanded to account for more choice features. Standard practices have not yet developed to measure out-of-pocket trip costs. More work is needed to integrate stated preference data into travel cost models.

Other loose ends are identified in Chap. 6. For example, Sect. 6.3.8 highlights the components of trip costs (travel costs, the value of travel time, equipment costs, and access fees) and points out that the travel cost method assumes that these costs are treated as given by subjects, whereas they may involve choices by subjects. As a case in point, consider the choice of where one lives, a determinant of travel expenses and time spent in travel. At its heart, the travel cost method uses the behavior of those with higher travel costs to predict participation rates of people with lower travel costs if the price of visits were raised. But what if people with different travel costs also have different preferences regarding the recreational activity in question? If some people choose where they live based in part on nearness to recreation sites, a serious bias could be introduced. Rock climbing, fly fishing, and downhill skiing come immediately to mind as examples. Other things being equal, the result would be to underestimate values.

As mentioned previously, still another issue is recall bias. If people report visiting sites sites more frequently than they actually, this will bias values.

These are not trivial issues. If any of these or other issues arise, the resulting values are likely biased. Such soft spots reduce the content validity of the travel cost method. This can be partially counterbalanced by carefully reporting assumptions used in travel cost studies. Robustness tests of critical assumption are also helpful.

12.4.2.2 Travel Cost Construct Validity

The construct validity of the travel cost method can be considered from three different angles. First, does it produce values that are economically plausible? Second, can it demonstrate relationships between value estimates and priors drawn from theory and intuition? And third, do travel cost values demonstrate convergence with values derived from other methods?

State-of-the-art travel cost studies normally produce value estimates that are intuitively reasonable or plausible, adding to the credibility of the method. Of course, this test has limited potency because estimated values could be plausible but still contain bias. However, if travel cost studies consistently produced outlandish value estimates—let us say thousands of dollars per day of recreation—researchers would have abandoned it long ago. Instead they have looked at study after study and concluded that, yes, the recreational activities under study could be worth what the analyses concluded. To the counter, we know of no literature that argues that these values are not plausible.

As for prior expectations, travel cost applications consistently find a negative relationship between travel costs and participation, satisfying negative price sensitivity.

Income sensitivity is more interesting. Do values from travel cost models show a fairly consistent tendency to increase with subjects' incomes? This question is not often asked. Many studies in the peer-reviewed literature since 2000—including Boxall and Adamowicz (2002), Lupi et al. (2003), Haener et al. (2004), Moeltner and Englin (2004), Kinnell et al. (2006), Hynes et al. (2007), Timmins and Murdock (2007), Scarpa et al. (2008), and Hindsley et al. (2011)—did not include income as a possible explanatory variable. Of the exceptions, some found positive, significant income sensitivity (Boxall et al. 2003; Landry and Liu 2009); others found no significant effect of income (Massey et al. 2006); and some had more than one model with mixed results (Grijalva et al. 2002; Murdock 2006). While it would be surprising if a thorough investigation failed to demonstrate positive income sensitivity, no such investigation appears to have been conducted. Construct validity has not been confirmed in this dimension.

The counterparts of the contingent valuation scope effect are the many travel cost studies—including several cited in Chap. 6—that value changes in the availability of recreation sites and/or the quality of those sites. These studies often find statistically significant values. See, for example, the beach case study in Sect. 6.4. This is positive evidence for the construct validity of the travel cost method.

Meta-analyses provide some support for the construct validity of the travel cost method by documenting that value estimates demonstrate expected relationships to many independent variables. For example, Smith and Kaoru (1990a) found that travel cost value estimates vary systematically with the type of recreation activity valued. Shrestha and Loomis (2003) found recreational values were related to whether the site was a lake, river, or forest; whether the site was publicly owned; whether the site had developed recreational facilities (picnic tables, campgrounds, etc.); the number of different kinds of recreational opportunities available at the site; and the type of recreational opportunity being valued, which ranged from camping to big-game hunting to snowmobiling. This supports the construct validity up to a point, but unfortunately the results are somewhat clouded by the fact that both travel cost and contingent valuation value estimates were included. Following similar procedures but using only travel cost values would be useful.

Regarding convergent validity, recall the discussion in the preceding section of tests comparing stated and revealed preference methods, including the travel cost method. While the contingent valuation discussion concluded that convergence has not been fully confirmed, travel cost and stated preference methods do provide some mutually reinforcing evidence of convergent validity.

12.4.2.3 Travel Cost Criterion Validity

Only one study appears to have compared travel cost values with simulated market values. McCollum (1986) compared travel cost estimates for a deer hunting experience with WTP estimates from simulated markets for the same deer hunting opportunity. The main finding was that behavior captured in travel cost models was not statistically distinguishable from the behavior in cash markets when relatively low values for travel time (10-33% of the wage rate) were used. Chapter 6 notes that the most common opportunity cost of travel time used in the literature is one-third of the wage rate. These results suggest that travel cost estimates of value might be valid for a one-time recreation activity when modest values are applied to the opportunity cost of travel time. More studies comparing travel cost values to simulated market values would be helpful.

12.4.3 Travel Cost Accuracy and the Weight of Evidence

What can be said by way of conclusions about the reliability and validity of the travel cost method?

Reliability may not be an issue if respondents are asked to recall trips over relatively short and recent periods. This is good news for the reliability of studies that can acquire data with short, recent recall. Still, more research is needed on the reliability of the travel cost method using test-retest methods.

On the validity side, consider each of the three legs of the validity stool.

While **content validity** of the travel cost method seems well established, concern remains about the assumptions that those applying the method must make without clear theoretical or empirical guidance. In the meantime, carefully following the steps laid down in Chap. 6, documenting the choices made, and conducting robustness tests can enhance content validity.

Regarding **construct validity**, several findings are encouraging: the plausibility of value estimates, the consistent finding of negative price sensitivity, successes in efforts to value quality changes and changes in recreation site availability, and successes in testing for significant associations between values and prior expectations support the construct validity of the method. However, positive income sensitivity remains to be demonstrated. Turning to convergent validity, travel cost studies tend to give higher values than stated preference studies. Given that the two approaches are so different and the value estimates are correlated, the closeness of the value estimates from the two methods is nevertheless encouraging.

There is not much evidence of **criterion validity**, pro or con, for the travel cost method. Existing literature suggests that relatively simple travel cost models using modest allowances for travel time are the most likely to be valid. More research is definitely needed and simulated markets comparisons are one promising line of attack.

In sum, the weak leg for criterion validity makes the three-legged stool for the travel cost method seem a bit shaky. One could speculate that the confidence economists have in revealed preference data may have lulled practitioners into complacency about fully investigating the validity of the travel cost method.

12.5 Conclusions

The accuracy framework is an outline to assist investigators in systematically enhancing the reliability and validity of nonmarket valuation studies. If you think of that framework as a skeleton, research to date has done much to flesh out the body, but the work is not finished. The debate over the accuracy of nonmarket valuation methods has been most intense when focused on contingent valuation. Much has been accomplished since Scott (1965) sarcastically wrote, "ask a hypothetical question and you get a hypothetical answer" (p. 37). The weight of evidence suggests that much progress has been made since the early applications, and much more has been done since the NOAA panel's pronouncements. The contingent valuation debate has been healthy even if it has been uncomfortable at times. Such rigorous debate needs to continue and expand to the other nonmarket valuation techniques.

Thinking about the evolution of nonmarket valuation, a disproportionately large share of the research appears to have focused on improving econometric estimation methods compared to other issues relating to study design and execution. Taking a more holistic perspective on accuracy from initial study conceptualization through value reporting, which this book has as a goal, can enhance the overall credibility of nonmarket value estimates. We believe that wide application of the accuracy framework introduced in this chapter should help to identify and balance research priorities for each of the valuation methods discussed in this book.

Criterion validity investigations are capable of providing the most potent tests of nonmarket valuation methods. More of this research is needed with specific attention given to controlled experiments where treatments are credible and appropriate weight is given to the evidence provided by each treatment in the experiments (i.e., not being unduly swayed by long-standing beliefs and biases of people defending one side or the other of the investigations).

A weakness of economics as a discipline compared to other disciplines is that, once a study is published, replications are very hard to get published; whereas in other disciplines, the publication of replication studies is a normal part of the scientific progress. The consequence of this disciplinary parochialism is that our base of knowledge may be broad, but it is not very deep, which undermines the ability to assess accuracy. This seems to be particularly true of nonmarket valuation.

Progress in improving the reliability and validity of nonmarket valuation methods has been slowed by lack of research funding, both from scientific funding agencies and from those who use nonmarket values. Thus, research is based on investigators' ability to cobble together funding to conduct small experiments or to attach their research to practical, policy-oriented studies for which funding is more generally available. This results in too few well funded studies specifically designed to address fundamental study design and implementation issues.

In spite of these limitations, the contingent valuation and travel cost examples illustrate the remarkable advances that have been made in the applications of stated and revealed preference methods. While acknowledging that there are investigator choices that need documentation and robustness checks, this is true for any empirical method, not just nonmarket valuation. Both examples have made great advances over the past 40+ years. For contingent valuation, an example is the use of incentive-compatible question formats in the context of consequential valuation exercises. For the travel cost method, this might be the use of random-utility models to more explicitly account for substitutes and effectively include resource quality in estimated models. Thus, returning to the glass of water metaphor, overall the nonmarket valuation "glasses" are at least half full and gaining volume (or accuracy). Our accuracy framework can enhance these gains when applied in a systematic fashion.

Acknowledgments We would like to thank Kerry Smith for his insightful comments that enhance this manuscript, and Vic Adamowicz, Rob Johnston, John Loomis, and George Parsons for providing suggested source material.

References

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. & Schuman, H. (1993). Natural resource damage assessments under the Oil Pollution Act of 1990. Federal Register, 58, 4601-4614.

Berrens, R. P., Bohara, A. K., Silva, C. L., Brookshire, D. & McKee, M. (2000). Contingent values for New Mexico instream flows: With tests of scope, group-size reminder and temporal reliability. Journal of Environmental Management, 58, 73-90.

Bjornstad, D., Cummings, R. & Osborne, L. (1997). A learning design for reducing hypothetical bias in the contingent valuation method. Environmental and Resource Economics, 10, 207-221.

Blinka, D. D. (2011). The *Daubert* standard in Wisconsin: A primer. Wisconsin Lawyer, 84 (3). Retrieved from www.wisbar.org/newspublications/wisconsinlawyer/pages/article.aspx?volume=84&issue=3&articleid=2348.

Boxall, P. C. & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: A latent class approach. Environmental and Resource Economics, 23, 421-446.

Boxall, P., Rollins, K. & Englin, J. (2003). Heterogeneous preferences for congestion during a wilderness experience. Resource and Energy Economics, 25, 177-195.

Brown, T. C., Ajzen, I. & Hrubes, D. (2003). Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation. Journal of Environmental Economics and Management, 46, 353-361.

Brown, W. G. & Nawas, F. (1973). Impact of aggregation on the estimation of outdoor recreation demand functions. American Journal of Agricultural Economics, 55, 246-249.

Burton, A. C., Carson, K. S., Chilton, S. M. & Hutchinson, W. G. (2007). Resolving questions about bias in real and hypothetical referenda. Environmental Resource Economics, 38, 513-525.
Carmines, E. G. & Zeller, R. A. (1979). Reliability and validity assessment. Newbury Park, CA: Sage Publications.
Carson, R. T. (1997). Contingent valuation surveys and tests of sensitivity to scope. In R. J. Kopp, W. W. Pommerehne, & N. Schwarz (Eds.), Determining the value of non-market goods (pp. 127-164). Norwell, MA: Kluwer Academic Publishers.
Carson, R. T., Flores, N.E., Martin, K. M. & Wright, J. L. (1996). Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods. Land Economics, 72, 80-99.
Carson, R. T., Flores, N. E. & Meade, N. F. (2001). Contingent valuation: Controversies and evidence. Environmental and Resource Economics, 19, 173-210.
Carson, R. T., Groves, T. & List, J. A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. Journal of the Association of Environmental and Resource Economists, 1, 171-207.
Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Rudd, P. A., Smith, V. K., Conaway, M. & Martin, K. (1997). Temporal reliability of estimates from contingent valuation. Land Economics, 73, 151-163.
Cesario, F. J. & Knetsch, J. L. (1970). Time bias in recreation benefit estimates. Water Resources Research, 6, 700-704.
Champ, P. A. & Brown, T. C. (1997). A comparison of contingent valuation and actual voting behavior. Proceedings, Benefits and Cost Transfer in Natural Resource Planning, 10th interim report, pp. 77-98.
Chu, A., Eisenhower, D., Hay, M., Morganstein, D., Neter, J. & Waksberg, J. (1992). Measuring the recall error in self-reported fishing and hunting activities. Journal of Official Statistics, 8, 19-39.
Cummings, R. G. & Taylor, L. O. (1998). Does realism matter in contingent valuation surveys? Land Economics, 74, 203-215.
Cummings, R. G., Elliott, S, Harrison, G. W. & Murphy, J. (1997). Are hypothetical referenda incentive compatible? The Journal of Political Economy, 105, 609-621.
Desvousges, W., Mathews, K. & Train, K. (2012). Adequate responsiveness to scope in contingent valuation. Ecological Economics, 84, 121-128.
Desvousges, W. H., Johnson, F. R., Dunford R. W., Boyle, K. J., Hudson, S. P. & Wilson, K. N. (1992). Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy. Research Triangle Park, NC: Research Triangle Institute.
Diamond, P. (1996). Testing the internal consistency of contingent valuation surveys. Journal of Environmental Economics and Management, 30, 337-347.
Diamond, P. A. & Hausman, J. A. (1993.) On contingent valuation measurement of nonuse values. In J. A. Hausman (Ed.), Contingent valuation: A critical assessment (pp. 3-38). Amsterdam: North Holland.
Diamond, P. A. & Hausman, J. A. (1994). Contingent valuation: Is some number better than no number? Journal of Economic Perspectives, 8 (4), 45-64.
Gems, B., Ghosh, D. & Hitlin, R. (1982). A recall experiment: Impact of time on recall of recreational fishing trips. Proceedings, Section on Survey Research Methods, 7, 168-173.
Grijalva, T. C., Berrens, R. P., Bohara, A. K., Jakus, P. M. & Shaw, W. D. (2002). Valuing the loss of rock climbing access in wilderness areas: A national-level, random-utility model. Land Economics, 78, 103-120.
Haab, T. C., Interis, M. G., Petrolia, D. R. & Whitehead, J. C. (2013). From hopeless to curious? Thoughts on Hausman's 'dubious to hopeless' critique of contingent valuation. Applied Economic Perspectives and Policy, 35, 593-612.
Haener, M. K., Boxall, P. C., Adamowicz, W. L. & Kuhnke, D. H. (2004). Aggregation bias in recreation site choice models: Resolving the resolution problem. Land Economics, 80, 561-574.

Hanley, N. D. (1989). Valuing rural recreation benefits: An empirical comparison of two approaches. Journal of Agricultural Economics, 40, 361-374.

Hausman, J. A. (Ed.). (1993). Contingent valuation: A critical assessment. Amsterdam: North Holland.

Hausman, J. A. (2012). Contingent valuation: From dubious to hopeless. Journal of Economic Perspectives, 26 (4), 43-56.

Heberlein, T. A., Wilson, M. A., Bishop, R. C. & Schaeffer. N. C. (2005). Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management, 50, 1-22.

Hindsley, P., Landry, C. E. & Gentner, B. (2011). Addressing onsite sampling in recreation site choice models. Journal of Environmental Economics and Management, 62, 95-110.

Hynes, S., Hanley, N. & Garvey, E. (2007). Up the proverbial creek without a paddle: Accounting for variable participant skill levels in recreational demand modelling. Environmental and Resource Economics, 36, 413-426.

Jacobsen, J. B. & Hanley, N. (2009). Are there income effects on global willingness to pay for biodiversity conservation? Environmental and Resource Economics, 43, 137-160.

Johnston, R. J. (2006). Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum. Journal of Environmental Economics and Management, 52, 469-481.

Jones-Lee, M., Hammerton, M. & Phillips, P. R. (1985). The value of safety: Results of a national survey. Economic Journal, 95, 49-72.

Kahneman, D. & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. Journal of Environmental Economics and Management, 22, 57-70.

Kealy, M. J., Dovidio, J. F. & Rockel, M. L. (1988). Accuracy in valuation is a matter of degree. Land Economics, 64, 158-171.

Kealy, M. J., Montgomery, M. & Dovidio, J. F. (1990). Reliability and predictive validity of contingent values: Does the nature of the good matter? Journal of Environmental Economics and Management, 19, 244-263.

Kinnell, J. C., Bingham, M. F., Mohamed, A. F., Desvousges, W. H., Kiler, T. B., Hastings, E. K. & Kuhns, K. T. (2006). Estimating site choice decisions for urban recreators. Land Economics, 82, 257-272.

Kochi, I., Hubbell, B. & Kramer, R. (2006). An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis. Environmental and Resource Economics, 34, 385-406.

Landry, C. E. & List, J. A. (2007). Using ex ante approaches to obtain credible signals for value in contingent markets: Evidence from the field. American Journal of Agricultural Economics, 89, 420-429.

Landry, C. E. & Liu, H. (2009). A semi-parametric estimator for revealed and stated preference data—An application to recreational beach visitation. Journal of Environmental Economics and Management, 57, 205-218.

List, J. A. & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics, 20, 241-254.

Little, J. & Berrens, R. (2004). Explaining disparities between actual and hypothetical stated values: Further investigation using meta-analysis. Economics Bulletin, 3 (6), 1-13.

Loehman, E. & De, V. H. (1982). Application of stochastic modeling to the study of public goods. Review of Economics and Statistics, 64, 474-480.

Loomis, J. B. (1989). Test-retest reliability of the contingent valuation method: A comparison of general population and visitor responses. American Journal of Agricultural Economics, 71, 76-84.

Loomis, J. B. (1990). Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques. Journal of Environmental Management, 18, 78-85.

Lupi, F., Hoehn, J. P. & Christie, G. C. (2003). Using an economic model of recreational fishing to evaluate the benefits of sea lamprey (*Petromyzon marinus*) control on the St. Marys River. Journal of Great Lakes Research, 29 (Supplement 1), 742-754.

Massey, D. M., Newbold, S. C. & Gentner, B. (2006). Valuing water quality changes using a bioeconomic model of a coastal recreational fishery. Journal of Environmental Economics and Management, 52 (1), 482-500.

Mazurkiewicz, S. M., Boyle, K. J., Teisl, M. F., Morris, K. I. & Clark, A. G. (1996). Recall bias and reliability of survey data: Moose hunting in Maine. Wildlife Society Bulletin, 24, 140-148.

McCollum, D. W. (1986). The travel cost method: Time, specification, and validity. (Doctoral dissertation). University of Wisconsin, Madison.

McConnell, K. E., Strand, I. E. & Valdés, S. (1998). Testing temporal reliability and carry-over effect: The role of correlated responses in test-retest reliability studies. Environmental and Resource Economics, 12, 357-374.

McFadden, D. (1994). Contingent valuation and social choice. American Journal of Agricultural Economics, 76, 689-708.

Mitchell, R. C. & Carson, R. T. (1989). Using surveys to value public goods: The contingent valuation method. Washington, D.C.: Resources for the Future.

Moeltner, K. & Englin, J. (2004). Choice behavior under time-variant quality: State dependence versus 'play-it-by-ear' in selecting ski resorts. Journal of Business and Economic Statistics, 22, 214-224.

Moore, J. C., Stinson, L. L. & Welniak, E. J. (2000). Income measurement error in surveys: A review. Journal of Official Statistics, 16, 331-362.

Murdock, J. (2006). Handling unobserved site characteristics in random utility models of recreation demand. Journal of Environmental Economics and Management, 51, 1-25.

Murphy, J. J., Allen, P. G., Stevens, T. H. & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics, 30, 313-325.

Reiling, S. D., Boyle, K. J., Cheng, H. & Philips, M. L. (1989). Contingent valuation of a public program to control black flies. Northeastern Journal of Agricultural and Resource Economics, 18, 126-134.

Reiling, S. D., Boyle, K. J., Phillips, M. L. & Anderson, M. W. (1990). Temporal reliability of contingent values. Land Economics, 66, 128-134.

Rosenberger, R. S. & Loomis, J. B. (2000). Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database. Water Resources Research, 36, 1097-1107.

Scarpa, R., Thiene, M. & Train, K. (2008). Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps. American Journal of Agricultural Economics, 90, 994-1010.

Schkade, D. A. &. Payne, J. W. (1994). How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation. Journal of Environmental Economics and Management, 26, 88-109.

Schläpfer, F. (2006). Survey protocol and income effects in the contingent valuation of public goods: A meta-analysis. Ecological Economics, 57, 415-429.

Scott, A. (1965). The valuation of game resources: Some theoretical aspects. Canadian Fisheries Report, 4, 27-47.

Shrestha, R. K. & Loomis, J. B. (2003). Meta-analytic benefit transfer of outdoor recreation economic values: Testing out-of-sample convergent validity. Environmental and Resource Economics, 25, 79-100.

Smith, V. K. & Kaoru, Y. (1990a). Signals or noise? Explaining the variation in recreation benefit estimates. American Journal of Agricultural Economics, 72, 419-433.

Smith, V. K. & Kaoru, Y. (1990b). What have we learned since Hotelling's letter? Economic Letters, 32, 267-272.

Stevens, J. B. (1969). Effects of nonprice variables upon participation in water-oriented outdoor recreation: Comment. American Journal of Agricultural Economics, 51, 192-193.

Stumborg, B. E., Baerenklau, K. A. & Bishop, R. C. (2001). Nonpoint source pollution and present values: A contingent valuation study of Lake Mendota. Review of Agricultural Economics, 23, 120-132.

Tarrant, M. A., Manfredo, M. J., Braley, P. B. & Hess, R. (1993). Effects of recall bias and nonresponse bias on self-report estimates and angling participation. North American Journal of Fisheries Management, 13, 217-222.

Taylor, L. O. (1998). Incentive compatible referenda and the valuation of environmental goods. Agricultural and Resource Economics Review, 27, 132-139.

Taylor, L. O., McKee, M., Laury, S. K. & Cummings, R. G. (2001). Induced-value tests of the referendum voting mechanism. Economics Letters, 71, 61-65.

Teisl, M. F., Boyle, K. J., McCollum, D. W. & Reiling, S. D. (1995). Test-retest reliability of contingent valuation with independent sample pretest and posttest control groups. American Journal of Agricultural Economics, 77, 613-619.

Timmins, C. & Murdock, J. (2007). A revealed preference approach to the measurement of congestion in travel cost models. Journal of Environmental Economics and management, 53, 230-249.

Trice, A. H. & Wood, S. E. (1958) Measurement of recreation benefits. Land Economics, 34, 195-207.

Vossler, C. A. & Kerkvliet, J. (2003). A criterion validity test of the contingent valuation method: Comparing hypothetical and actual voting behavior for a public referendum. Journal of Environmental Economics and Management, 45, 631-649.

Vossler, C. A., Kerkvliet, J., Polasky, S. & Gainutdinova, O. (2003). Externally validating contingent valuation: An open-space survey and referendum in Corvallis, Oregon. Journal of Economic Behavior & Organization, 51, 261-277.

Zeller, R. A. & Carmines, E. G. (1980). Measurement in the social sciences: The link between theory and data. New York: Cambridge University Press.

Index

P.A. Champ et al. (eds.), *A Primer on Nonmarket Valuation*,
The Economics of Non-Market Goods and Resources 13,
DOI 10.1007/978-94-007-7104-8