Chapter 13

PERFORMANCE OF ERROR COMPONENT MODELS FOR STATUS-QUO EFFECTS IN CHOICE EXPERIMENTS

Riccardo Scarpa*
Environment Department
University of York
York, U.K.
rs24@york.ac.uk

Silvia Ferrini
Department of Statistics
University of Florence, Florence, Italy
ferrini@ds.unifi.it

Kenneth Willis
School of Architecture and Planning
University of Newcastle University
Newcastle upon Tyne, U.K.
kenneth.willis@newcastle.ac.uk

Abstract Environmental economists have advocated the use of choice modelling in environmental valuation. Standard approaches employ choice sets including one alternative depicting the status-quo, yet the effects of explicitly accounting for systematic differences in preferences for non status-quo alternatives in the econometric models are not well understood. We explore three different ways of addressing such systematic differences using data from two choice modelling

*We acknowledge constructive comments on previous drafts by Giovanni Baiocchi, Elisabetta Cherchi, Joseph Herriges, Stephane Hess, Andreas Kontoleon, Luis Rizzi, two anonymous reviewers and participants in the Budapest 2004 EAERE meeting. All remaining errors are entirely our responsibility.

studies designed to value the provision of environmental goods. Preferences for change versus status-quo are explored with standard conditional logit with alternative-specific constant for status-quo, nested logit and a less usual mixed logit error component specification (kernel logit). Our empirical results are consistent with the hypothesis that alternatives offering changes from status-quo do not share the same preference structure as status-quo alternatives, as found by others in the marketing literature, in the environmental economic literature and in food preference studies. To further explore the empirical consequences of such mis-specification we report on a series of Monte Carlo experiments. Evidence from the experiments indicates that the expected bias in estimates ignoring the status-quo effect is substantial, and—more interestingly—that error component specifications with status-quo alternative specific-constant are efficient even when biased. These findings have significant implications for practitioners and their stance towards the strategies for the econometric analysis of choice modelling data for the purpose of valuation.

Keywords: choice-modelling, stated-preference, environmental valuation, status-quo bias, Monte Carlo simulations, water resources.

1. Introduction

Since their early appearance in the environmental economics literature in the middle-to-late nineties (Roe *et al.* 1996; Boxall *et al.*, 1996; Garrod and Willis, 1997; Adamowicz *et al.*, 1998) "choice experiments"[1] have enriched and further diversified the non-market valuation applications based on stated preferences. The number of studies on this methodology has been rapidly growing (Layton 2000; Morrison *et al.*, 2002; Foster and Mourato, 2002; Garrod *et al.* 2002) with applications covering many non-market valuation contexts. Overall, the role of this approach in diversifying the field of non-market valuation has been eloquently praised (Randall, 1998).

The basic method requires respondents to indicate a preference ordering — by ranking, rating or identifying a preferred choice — over a set of experimentally designed alternatives. Although the inclusion or exclusion of the status-quo (henceforth abbreviated in SQ) in the choice-set depends on the objective of the survey (see Breffle and Rowe, 2002 for a discussion), to increase realism (Ortúzar and Willumsen, 2001) most studies in transportation and environmental economics are based on survey designs that include a SQ alternative. This is often described to respondents in terms of the attribute values that are experienced and associated with the SQ. An issue that remains little explored to date is whether or not respondents "perceive"—and as a consequence evaluate—the alternatives associated with change from the SQ somewhat differently from the SQ alternative. This asymmetry would be consistent with reference-dependent

[1] Or "choice modelling", or "conjoint choice analysis" or more generally multi-attribute stated preferences.

utility theories (Kahneman and Tversky, 1979; Hartman *et al.* 1991; Samuelson and Zeckhauser, 1988; Bateman *et al.* 1997). This is a key issue because if such a difference exists, then it raises, amongst others, three problems relevant for choice experiments. First, how should one explicitly account for this effect in the analysis? Second, what are the consequences of accounting for it in an incorrect fashion, that is, the consequences of mis-specification? Third, what difference should one expect when using different sample sizes within the conventional range?

We find that these issues have yet to be satisfactorily addressed in the multi-attribute stated preference literature, and so, the main objective of this chapter is to explore such problems focusing on the finite sample properties of welfare estimates using experiments based on empirical results.

We compare three different ways of modelling diversity in perception of SQ versus alternatives involving change. All three are based on the conventional random utility framework.

In principle, one can argue that there are two kind of effects when a SQ alternative is present in all choice-sets. The first is a systematic effect which can be easily estimated by means of an alternative-specific SQ constant in the utility function. The second is an effect on the stochastic error structure postulated by the researcher. For example, designed alternatives involving change from the SQ one can share an error structure with a stochastic behaviour that is more similar to each other than it is to the error associated with the SQ alternative. In other words, designed alternatives involving change are correlated, and their error component is hence not independent. Such correlation can be accommodated within a nested logit framework.

In the literature SQ effects are typically dealt with by two specifications: the conditional logit with alternative-specific SQ constants and the nested logit. The first addresses systematic SQ effects, the second the correlation across utilities of designed alternatives. Both have been tested and found statistically significant. In this paper we propose a third specification, which flexibly and simultaneously addresses both types of effects by means of an error component mixed logit specification with alternative-specific SQ constant. This flexible model induces a correlation pattern in the utility of alternatives involving change, as well as capturing a systematic effect due to the SQ in the indirect utility.[2]. In one of the two datasets employed here is found to be a significant improvement over its competitors.

As a backdrop to such an investigation we report on findings from a large-scale survey designed to value the public good provision associated with water supply to the residents of the counties of Yorkshire, in the U.K. The analysis

[2]We are grateful to Joseph Herriges for suggesting this last specification during the 2004 EAERE meeting in Budapest, where a previous version of this paper was presented.

of these data provide evidence in support of the hypothesis of a systematic difference in customers' evaluations of SQ and non-SQ alternatives across the three specifications.

The finite sample properties of each of these three estimators are then analyzed using Monte Carlo experiments. These are conducted at three sample sizes that cover the range most commonly employed in the literature. The results of these experiments provide valuable information on the potential size of the bias in the estimates in the presence of SQ effects. They also provide suggestions on the relative efficiency of various estimators under misspecification.

The remainder of the paper is organized as follows. The next section briefly outlines the motivations for the investigation. Section 3 defines a common notation for the various models. In section 4 we describe the studies from which we draw inspiration for the Monte Carlo experiment. The results of the studies are presented in section 5, while the evidence from the Monte Carlo experiments is discussed in section 6. In section 7 we conclude.

2. Motivation for the study

2.1 The nature of SQ effects

For the purpose of this paper we define "SQ effects" as the systematic inclination of respondents to display a different attitude towards SQ alternatives from those reserved to alternatives involving some change, over and beyond what can be captured by the variation of attributes' levels across alternatives.

In the context of public economics, of which environmental economics is a sub-discipline, we are often concerned with a SQ resulting from previous policy outcomes, and incorporating public views on property rights, institutional arrangements etc. This extends to endowment of passive use and non-use values, which overall represent a bundle of issues conceptually quite different from those embedded in the SQ alternative employed in transportation and market research choice experiments, where the emphasis is on *use* values.[3]

However, regardless of its prevailing nature in environmental economics, this effect seems quite general. In their much quoted paper on consumer rationality and SQ effects (or "bias") Hartman *et al.* (1991) write:

> This analysis suggests, for example, that consumers attach "undue" importance to their current commodity bundle, demonstrating "apparently irrational" reluctance to alternative bundles. (Page 141)

The explanatory nature of such a general phenomenon is quite complex. For example, in a previous contribution on the topic Samuelson and Zeckhauser

[3]In transportation the tendency to systematically prefer the SQ alternative is termed "attrition", e.g. Bradley and Daly, 1997; Cantillo and Ortúzar, 2004.

(1988) identified and validated with evidence three major categories of explanations for this type of behaviour:

1. rational decision making in the presence of transition costs and/or uncertainty;
2. cognitive mis-perceptions (e.g. loss aversion and prospect theory as proposed by Kahneman and Tversky (1979))
3. psychological commitment stemming from misperceived sunk costs, regret avoidance, or a drive for consistency.

In their conclusions from the study Samuelson and Zeckhauser write:

> In choosing among alternatives individuals display a bias toward sticking with the status quo.[...] Assuming the status quo bias proves important, rational models will present excessively radical conclusions, exaggerating individual's responses to changing economic variables and predicting greater probability than observed in the world.

In our experience with SQ effects in choice experiments such effects can show both, a predilection for the SQ or a reluctance to stick with it.[4]

Because of this multiple causes of SQ bias, we do not find fruitful to elaborate on a conceptual model, which inevitably will leave some explanations unaccounted for. Hence, in what follows we maintain the conceptual definition quite general, yet we specifically focus on the analysis of multinomial discrete choices for environmental valuation under a random utility framework.

2.2 SQ effects in choice experiments

Above we have referred to some evidence from psychology and experimental economics suggestive that people evaluate what they know, and are familiar with (i.e. the SQ), in a systematically different fashion from how they evaluate hypothetical alternative scenarios. This has a direct bearing in the application of choice experiments to non-market valuation of public goods. Whether the technique is used to expand the set of modelling approaches existing in practice (as is often argued in support of mixing stated and revealed preference data (e.g. Hensher *et al.*, 1999), the so called "data fusion" approach), or is employed as a way to elicit trade-offs that will eventually lead to a richer description of people's preference for environmental goods (e.g. when respondents are asked to evaluate SQ scenarios against hypothetical changes), a check for what we loosely call "SQ effects" should be performed. The specification that best addresses such an effect will depend on circumstances.

[4] For further reference to status quo effects in power outages discrete choice elicitation studies see section 2 in this volume.

If a "generic" SQ effect exists in the evaluation of alternatives of choice-experiments for non-market goods, then an adequate practical understanding of the various econometric approaches accounting for such bias needs to be developed. By investigating the finite sample properties of common estimators we contribute to the on-going research on the understanding of the implications of modelling choices for the derivation of welfare estimates from choice-experiments investigating non-market goods.[5]

In particular, a systematic investigation of standard RUM-based modelling approaches to SQ effects in non-market valuation seems to be missing, and this is what we set out to provide here.

Some work on this issue is to be found in market research (Haaijer 1999; Haaijer *et al.* 2001), but this is limited to comparing Nested logit and conditional (or multinomial)[6] logit with alternative-specific constant for the SQ. Furthermore, the study is prevalently concerned with technical aspects proper of market research (coding effects, brand effects, market shares etc.). Their results suggest that the violation of the independence of irrelevant alternative makes the use of models not reliant on such restriction appealing, an argument that is often used in promoting the use of random parameter specifications for the indirect utility (Layton, 2000; Garrod *et al.*, 2002; Kontoleon and Yobe, 2003). This is a suggestion that we explore in more detail here by using more flexible models, but focussing on a basic error components specification, rather than on one with random parameters.

In practice we report the results of an investigation comparing multinomial logit with an alternative specific constant for the SQ (MNL-Asc), nested logit (NL), and an error component mixed logit (MXL-ε) which also includes an an alternative specific constant for the SQ. These encompass those models to which most practitioners would turn to, at least in the first instance, when trying to account for SQ effects in econometric specifications. In general our results show that the conventional practice of using simply a MNL-Asc may often be unsound. In particular, the proposed error component model with alternative-specific constant for the SQ, that would appear to be novel in this literature, seems to perform better than others in most of the circumstances examined here.

In the following we focus on the case in which the choice set contains only three alternatives: the SQ and two other alternatives, all of which are described

[5]For example, an original avenue of investigation based on hurdle models and more broadly centred on non-participation is to be found in von Haefen and Adamovicz (2004). Their results, like ours, show the sensitivity of welfare estimates to the treatment of SQ and non-participation choices.

[6]Somewhat loosely, the terms *conditional* and *multinomial* logit are used as synonymous in this chapter.

on the basis of attribute levels, and from which respondents are asked to select the one they prefer.[7]

Although there are many variations on the theme, the prevalent set-up for choice experiments in environmental economics tends to present respondents with a choice task that involves the identification of a preferred alternative from a choice set including the SQ and often two (e.g. Boxall *et al.*, 1996; Hanley *et al.*, 1998; Rolfe *et al.*, 2000; Foster and Mourato, 2003; Scarpa *et al.*, 2003; Lehtonen *et al.*, 2003) or sometimes few more (e.g. Kontoleon and Yabe, 2003) experimentally designed alternatives. This set-up is often argued on the basis of a lower cognitive burden on respondents than that associated with other choice contexts in which the complexity of choice task is higher. In experiments involving ranking – for example – especially with many experimentally designed alternatives, it has long been noticed that the hypothesis of identical preference across decisions at different ranks is empirically violated (Hausman and Ruud 1987; Ben-Akiva *et al.* 1992).

3. Econometric specifications accounting for SQ and their rationale

3.1 Common notation

It is useful to start by defining a common notation for the various models, referring as much as one can to convention. The reference structure is the case where the analytical objective is to obtain maximum likelihood estimates of a $1 \times k$ row vector of utility weights $\boldsymbol{\beta}$ for a column vector $\mathbf{x}$ of $k \times 1$ attributes for the individual linear indirect utility function V_j. The available data are choices from choice tasks including a SQ (indexed as sq) and a minimum of two experimentally designed alternatives (indexed with subscripts c_1, c_2). This basic implementation is often encountered in the published literature, but it can be extended without loss of generality. For the purpose of valuation, welfare estimates can be obtained as (non-linear) functions of the estimates, using the usual difference between the logsums weighted by the inverse of the cost coefficient.

3.2 Conditional logits with and without SQ-Asc

The basic random utility consistent model for analyzing choice experiment data is the conditional logit, which we consider here as a baseline.

When u_{nj}, the stochastic component of utility for respondent n and alternative j, is identically and independently Gumbel distributed across all alterna-

[7] The main results of our study can, however, be generalised to many other contexts in which respondents are asked to rank any set of alternatives including the one depicting the SQ.

tives, then the choice probability is expressed by the well-known formula:

$$P_n(i) = \frac{\exp(\lambda V_i)}{\sum_j \exp(\lambda V_j)},\ j = sq, c_1, c_2 \tag{3.1}$$

where λ is the scale parameter of the unobserved stochastic component. This is the conventional conditional logit model, which we refer to as MNL. Here any diversity in preference for the alternatives different from the SQ may be explicitly made part of the non-stochastic component of utility, for example in the form of an alternative specific constant (Asc) which takes the form of an indicator function $Asc = 1$ iff $j = sq$.

For a simple example of a positive value on the SQ Asc consider the tendency for respondents who perceive the cognitive task of assessing all the alternatives as too daunting, to fall back on the familiar SQ, rather than engaging into a costly and unrewarded cognitive task.

In practice, significance of Asc parameter represents the most immediate SQ test. Although often the ASCs are associated with the designed alternatives, and the SQ is left as a baseline (Hanley and Wright 2003), we prefer here to use a dummy variable for the alternative describing the SQ, rather than one for each of the alternatives involving change, as advocated in Adamovicz *et al.* (1998). This specification allows the analyst to account for diversity in the probability of choice between hypothetical alternatives and experienced SQ. Notice, however, that this solution does not change the stochastic structure of the model as it only enters the *deterministic* component of utility, leaving the *stochastic* error structure unchanged. As such it does not allow for a varying correlation structure across alternatives, which instead we find to be quite a plausible hypothesis in behavioural terms.

Under linear indirect utility $V_j = \boldsymbol{\beta}'\mathbf{x}_j$ our parsimonious specification is therefore:

$$P(c_1) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}}}{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})} + e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.2}$$

$$P(c_2) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}}{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})} + e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.3}$$

$$P(sq) = \frac{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})}}{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})} + e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.4}$$

where—for convenience—the subscript n denoting individuals is ignored, and the scale parameter λ is standardized to 1 and hence it is omitted.

Note that if hypothetical changes are expected to increase utility, then the sign of Asc will be negative, and positive if the effect has the opposite direction. This is the alternative specific constant conditional logit model, which we refer to as MNL-Asc.

3.3 Nested logit

The different nature of the SQ alternative vis-à-vis the two experimentally designed ones may translate into a difference in the substitution patterns, and hence in a different correlation structure of the unobserved components of the individual utilities. One assumption consistent with this case is when the stochastic component of utility is distributed according to a generalized extreme value (GEV) distribution, then different patterns of correlation across the utility of alternatives can be generated, although these are subject to considerable restrictions (Train, 2003). In fact, correlations are imposed to be similar within nests, but for alternatives in different nests the unobserved components are uncorrelated, and indeed independent. This is the case of the nested logit in which the unobserved components of utility have the GEV cumulative distribution:

$$\exp\left[-\sum_{g=1}^{G}\left(\sum_{j\in J}\exp(-u_{ig}/\eta_g)\right)^{\eta_g}\right] \tag{3.5}$$

where g denotes nests. In the set-up we consider here, with one SQ alternative and two experimentally designed ones, the assumption that the correlation amongst unobserved stochastic components differs between the two sets of alternatives generates two nests. The first is a degenerate one associated with the SQ alternative. The second is associated with changes from the SQ and contains both the experimentally designed ones. This gives rise to the following probability structure for the first decision stage:

$$P(change) = \frac{e^{\eta VI}}{e^{\eta VI} + e^{\boldsymbol{\beta}'\mathbf{x}_{sq}}} \text{ and } P(sq) = 1 - P(change)$$

While for the second decision stage, which is given the decision of embracing some change, is:

$$P(c_1|change) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}}}{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \text{ and } P(c_2|change) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}}{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.6}$$

$$P(c_j) = P(change)P(c_j|change) = \frac{e^{\eta VI}}{e^{\eta VI} + e^{\boldsymbol{\beta}\mathbf{x}_{sq}}} \times \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_j}}}{e^{\boldsymbol{\beta}\mathbf{x}_{c_j}} + e^{\boldsymbol{\beta}\mathbf{x}_{c\neq j}}} \tag{3.7}$$

where $VI = \ln\left[\exp\left(\boldsymbol{\beta}\mathbf{x}_{c_1}\right) + \exp\left(\boldsymbol{\beta}\mathbf{x}_{c_2}\right)\right]$ and can be interpreted as a measure of the expected utility of accessing the nest with the alternatives associated with change. The reader is reminded that the coefficient η is a measure of dissimilarity between alternatives in the various nests, while the value $1-\eta$ is a proxy for correlation for alternatives within the same nest. In this context a higher value of η can be intuitively interpreted as higher utility weight of moving away from the SQ.

A number of recent choice experiment studies in environmental economics have used nested logit models to account for SQ effects and found them superior in terms of fit to their MNL counterparts (Blamey *et al.* 2002; Hanley and Wright, 2003; Lehtonen *et al.*, 2003; Li *et al.*, 2004).

It is noteworthy that although this model maintains the independence of irrelevant alternatives (IIA) property across alternatives belonging to the same nest, it allows for differences in cross-elasticities across nests.

3.4 Error components via mixed logit

Notice that neither the MNL-Asc nor the NL specifications simultaneously identify both the *systematic* and *stochastic* components of the SQ effect, nor do they allow for taste-heterogeneity, or break completely[8] away from the IIA assumption. A specification that may overcome all these limitations is the mixed logit with error components. It does so by allowing flexible patterns of substitution via an induced correlation structure across utilities.

This is, of course, a special case of the large family of mixed logit, which—as described in McFadden and Train (2002)—with adequate data quality, may in principle be used to approximate any type of RUM.

The richness and flexibility of mixed logit models have been shown to generate a large variety of correlation patterns (Brownstone and Train, 1999; Train, 2003; Munizaga and Alvarez, 2001; Herriges and Phaneuf, 2002). Train (2003, page 156) discusses eloquently how mixed logit can give rise to two quite different interpretations, the random parameter and, under some restrictions, the error component one (or kernel logit (Ben-Akiva *et al.* 2001)). Further considerations, more specific to transportation applications, can be found in Cherchi and Ortúzar (2004), and some potential drawbacks are discussed in Hensher and Greene (2003).

Specifications using random utility parameters are well-known and often employed in choice experiments designed for the valuation of environmental goods in their panel form, so as to account for repeated choices, break away from the IIA assumption and address unobserved heterogeneity. However, in our study we wish to maintain comparability across the underlying assumptions of the MNL-Asc and NL specifications, which do not allow taste-heterogeneity. We hence focus on the decomposition of the *unobservable* component of utility, rather than on random effects in the indirect utility, and adopt only an error component interpretation, something that is less frequently seen in this kind of literature. We exploit the fact that the inclusion of additional zero-mean error components in the structure of utility of each nest induces correlation patterns (Herriges and Phaneuf, 2002). In the presence of SQ effects

[8]The nested logit maintains it within each nest.

different correlation patterns exists between the unobservable components of utility of the SQ alternative, and those in alternatives involving change.

For example, in our choice experiment the error component approach takes the following basic utility form[9]:

$$\begin{aligned} U(c_1) &= \boldsymbol{\beta}\mathbf{x}_{c_1} + \tilde{u}_{c_1} = \boldsymbol{\beta}\mathbf{x}_{c_1} + \varepsilon_{c_1} + u_{c_1}, \\ U(c_2) &= \boldsymbol{\beta}\mathbf{x}_{c_2} + \tilde{u}_{c_2} = \boldsymbol{\beta}\mathbf{x}_{c_2} + \varepsilon_{c_2} + u_{c_2}, \\ U(sq) &= Asc + \boldsymbol{\beta}\mathbf{x}_{sq} + u_{sq} \end{aligned} \tag{3.8}$$

where, in our case, $\varepsilon_{c_1} = \varepsilon_{c_2} \sim N(0, \sigma^2)$ are additional error components to u_{c_1} and u_{c_2}, which are Gumbel-distributed with variance $\pi^2/6$, thereby leading to the following error covariance structure :

$$Cov(\tilde{u}_{c_1}, \tilde{u}_{c_2}) = \sigma^2, \;\; Var(\tilde{u}_{c_1}, \tilde{u}_{c_2}) = \sigma^2 + \pi^2/6, \tag{3.9}$$

$$Cov(\tilde{u}_{c_j}, \tilde{u}_{sq}) = 0, \;\; Var(\tilde{u}_{c_j}, \tilde{u}_{sq}) = \pi^2/6, \;\; j = 1, 2; \tag{3.10}$$

where $\tilde{u}_{c_j} = \varepsilon_{c_j} + u_{c_j}$. Note that this is an analog of the nested logit model in the sense that it allows for correlation of utilities across alternatives in the same nest, but different correlation for those across nests. However, there is no IIA restriction, and the Asc captures any remaining systematic effect on the SQ alternative.

Conditional on the presence of the error component ε_j the choice probability is logit, and the assumption above leads to the following expression for each marginal choice probability:

$$\begin{aligned} P(i) &= \int_{\varepsilon} P(i|\varepsilon) f(\varepsilon|\boldsymbol{\theta}) d\varepsilon \\ P(i) &= \int_{-\infty}^{+\infty} \frac{e^{\boldsymbol{\beta}\mathbf{x}_i + \varepsilon_i}}{\sum_j e^{\boldsymbol{\beta}\mathbf{x}_j + \varepsilon_j}} \phi(0, \sigma^2) d\varepsilon, \;\; j = c_1, c_2, sq \end{aligned} \tag{3.11}$$

where $\phi(\cdot)$ is the normal density, and $\varepsilon_j = 0$ when $j = sq$.

Notice, however, that the additional error component can be either independent across choices (for example in a non-panel structure) or it can be the same for all choices made by the same individual (in a panel implementation). This is relevant in choice experiments as it breaks away from assumption of independence in the error structure across choices by the same respondent, which is implicit in both conditional and nested logit assumptions. While in random parameter specifications it is more plausible to assume fixity of parameter across choices by the same respondent by means of panel estimators, it is less clear

[9]In fact, as expanded upon by Brownstone and Train (1999) and Herriges and Panheuf (2002), more general forms than this may be empirically appealing.

that this is the case for error components. Ultimately this remains an empirical issue to be assessed case-by-case, and within the same category of model and estimation procedure it can be assessed on the basis of log-likelihood values. In this paper we focus on a non-panel application and hence dwell on the conventional assumption of independence of error across choices by the same respondent. This assumption ensures comparability of the error component results with those from the other two specifications which also implies independence across choices.

In what follows, we refer to this error component mixed logit model with Asc as MXL-ε.

It is important to note that such model nests the other three, in the sense that a restriction of $\sigma = 0$ is consistent with the MNL-Asc model, a restriction of $Asc = 0$ is consistent with an analog of the NL model. Both restrictions return the MNL model.

3.5 Estimation issues

All models are estimated in GAUSS 3.6 by maximum likelihood methods, except for equation (3.11), which is estimated by maximum simulated likelihood (MSL) with Halton draws[10] (Train, 2000, 2003). The choice probability for alternative i is approximated by:

$$P(i) \approx \tilde{P}(i) = \frac{1}{S}\sum_{s=1}^{S} \frac{e^{\boldsymbol{\beta}\mathbf{x}_i + \varepsilon_i^s}}{\sum_j e^{\boldsymbol{\beta}\mathbf{x}_j + \varepsilon_j^s}} \tag{3.12}$$

where $\varepsilon_j^s = 0$ when $j = sq$, and s denotes simulation draws.

4. The Yorkshire Water choice modelling studies

In spring and summer 2002, as a part of a large-scale investigation into the preference structure of its customers, Yorkshire Water (YW) conducted a set of choice experiments. The aim was to characterize the preference for fifteen different attributes related to water provision, called here service factors (SFs). As a result of focus-group activities and discussion with the management, these SFs were separated into five groups, giving rise to five separate choice experiments. The first three were mostly concerned with SFs of a private good nature, and are ignored here.[11] In this chapter we are concerned with

[10]Model estimates were found to be stable at 50 Halton draws and obtained by using the GAUSS code made available by Kenneth Train. However, error component models can be estimates also in Nlogit by formulating adequate dummy variables and using the subcommand "dummy(n,*,0)" which restricts the mean of normally distributed parameter to be equal to zero.

[11]For a more extensive report the interested reader is referred to Willis and Scarpa, 2002 or Willis *et al.* 2004.

the two choice experiments that addressed attributes of the service that can be commonly interpreted as 'public goods'.

4.1 Study 1

The first choice experiment, defined here as 'study 1', looked at four service factors as attributes: area flooding by sewage (AF); river quality (RQ); nuisance from odour and flies (OF); and cost of service (change in water bill payment). There were eight levels of payment expressed as either increases or decreases on the current bill, while all other attributes were expressed at four levels as reported in Table 13.1. The design chosen was an orthogonal main effect factorial with a total of 32 profiles, which were split into sequences of four choices for each respondent. The design was obtained using SAS (for a survey of experimental designs for logit models using SAS see Kuhfeld, 2004).

The expected signs for the coefficient estimates were as follows. The percent of areas protected from sewage escape is indicated as AF (area flooding) and it is expected to show a positive sign. The percent of river length capable of supporting healthy fisheries is indicated as RQ (river quality) and it is also expected to show a positive sign. Finally, the number of households and business affected by odour and flies (OF) is expected to show a negative sign. Notice that this is more a club good than a public good, but it certainly has public good characteristics.

4.2 Study 2

The second choice experiment looked at three service factors as attributes: water amenities for recreation (AM) expected to be positive, quality of bathing water (BB) also expected to be positive, and cost of service obviously expected to be negative. There were seven levels of payment always expressed as increases on the current bill, while all other attributes were expressed at three levels (Table 13.1). The orthogonal main effect factorial design was obtained with SAS and gave a total of 27 cards, which were also split in sequences of four choices for each respondent.

4.3 Sampling method

The survey instrument was tested in a pilot study and further refined as a consequence. It was administered in person, by enumerators experienced with stated-preference questionnaires through a computer-assisted survey instrument. A representative sample of 767 Yorkshire Water residential customers completed the sequence of choices in study 1, for a total of 2,921 choices. A representative sample of 777 residential customers completed the sequence for study 2 experiment with a total of 3,108 choices. More detailed information

Table 13.1. Service levels (SF) for residential and business customers.

Abbreviation	*Factor*	*Description*	*Scaling*	*levels* : –1	*levels* : 0	*levels* : +1	*levels* : +2
AF	Sewage escape to land	% of areas protected from sewage escape in gardens, roads, paths and open areas	1	Coded =20%	Coded =35%	Coded =50%	Coded =100%
RQ	Ecological quality of rivers	% of river length capable of supporting healthy fisheries and other aquatic life in the long term	1	Coded =60%	Coded =75%	Coded =85%	Coded =90%
OF	Odour and flies	Number of households and businesses affected by odour and high numbers of flies from sewage treatment works	0.01	2000 Coded =20	600 Coded =60	300 Coded =30	150 Coded =15
AM	Ability to use inland waters for recreational use	Number of areas with waste water discharges designed to allow recreational activities on rivers	1		0	4	12
BB	Bathing beaches water quality	Sewage works and disinfections designed to meet government standards for bathing water	1		Meets current existing gov'nt standards coded =100%	Improvement: 50% better than gov'nt standard coded =150%	Improvement: 100% better than gov'nt standard coded= 200%

on the sampling methodology and the samples employed is available from the report to the water company (Willis and Scarpa, 2002).

5. Results and discussion

5.1 Estimates for study 1

The estimates for study 1 are reported in Table 13.2. Notice that the utility weights all have the anticipated signs for the attributes of the alternatives, and are statistically significant in all models. The inclusive value estimate in the nested logit model is in the (0-1] interval, and hence is consistent with utility maximization. The estimated spread of the error component (σ) is virtually zero in the MXL-ε model, which basically is equivalent to the MNL-Asc and NL models.

Table 13.2. Estimates for study 1, SFs: AF, RQ, OF. N = 2,921.

Coefficient	MNL	MNL-Asc	NL	MXL-ε
AF	0.011 (10.1)	0.017 (13.5)	0.018 (13.3)	0.017 (13.5)
RQ	0.057 (17.5)	0.070 (19.0)	0.075 (19.3)	0.07 (19.0)
OF	–0.125 (–19.91)	–0.130 (–18.8)	–0.137 (–18.4)	–0.130 (–18.8)
Cost	–0.159 (–25.7)	–0.135 (–20.5)	–0.142 (–20.7)	–0.135 (–20.5)
σ_ε				0.040 (0.121)
η			0.899 (112.9)	
SQ-Asc		0.604 (10.7)		0.604 (10.7)
MRS_{AF}	0.07 (0.06,0.08)	0.13 (0.10,0.15)	0.13 (–0.16,0.36)	0.13 (0.10,0.15)
MRS_{RQ}	0.36 (0.32,0.39)	0.52 (0.46,0.59)	0.53 (0.06,0.08)	0.52 (0.46,0.59)
MRS_{OF}	–0.79 (–0.87,–0.71)	–0.96 (–1.10 ,–0.85)	–0.96 (–1.17,–0.09)	–0.96 (–1.10,–0.085)
ln-L or ln-Sim$L$	–2,245	–2,185	–2,185	–2,185
AIC	4,498	4,500	4,500	4,502

Confidence intervals around marginal rates of substitution obtained with Krinsky and Robb (1986) method.

We observe that all three models accounting for SQ achieve a very similar fit according to the Akaike Information Criteria (AIC=-2ln$L$+2p).[12] The lowest log-likelihood is fitted by the conventional conditional logit.

The evidence is consistent with the hypothesis that there is a systematic and significant difference in perception and substitutability between *experimentally designed* alternatives and *experienced* SQ.

From the viewpoint of policy evaluation it is clear that customers of YW, feel strongly for the public goods associated with various water provision strategies. For example, the implicit *WTP* for a one percent increase in the area protected from sewage escape is valued by the average household between 0.07 and 0.13 pence.

Relatively more valuable is the percent increase in the length of river capable of supporting long-term fisheries, which gives a value ranging from 0.36 to 0.53 pence. A reduction of one hundred properties suffering nuisance from odour and flies is valued between 0.79 and 0.96 pence if we consider the point estimates across specifications.

In this sample, it is evident that choosing estimates that account for SQ bias in some form, does make a substantial difference, as the MNL model provides lower estimates than the other three models. However, within those accounting for SQ effects, the welfare estimates are of similar magnitude, with the exception of the NL estimates for AF.

5.2 Estimates for study 2

A similar pattern of considerations can be made for study 2 in which the experimentally designed alternatives never allowed for a decrease in public good provision, something that – instead – was allowed for in study 1, and that undoubtedly may increase the likelihood of violation of the IIA property.

Again, the estimates are consistent with the hypothesis that respondents perceived experimentally designed alternatives and SQ *differently*. Neglecting this fact would lead the analyst to infer lower *WTP* estimates for the public goods examined and to select models that were significantly worse in terms of AIC value. In study 2 (Table 13.3), however, there seems to be support for the hypothesis that the difference in perception between SQ and change should be incorporated in the *stochastic* component of utility, rather than in the systematic one. In fact, allowing for different correlation patterns (NL) improves the AIC by a much larger amount than allowing only for a systematic SQ effect in the deterministic component of utility (MNL-Asc).

[12]This criterion can be used to discriminate between un-nested models by placing a penalty on the number of parameters p, since NL is nested neither in the MNL nor in the MXL-ε. The model associated with the minimum value is to be considered the best (Akaike, 1973).

Table 13.3. Estimates for study 2, SFs: AM, BB. N = 3,180.

Coefficient	MNL	MNL-Asc	NL	MXL-ε
AM	0.067 (13.0)	0.079 (13.4)	0.095 (13.8)	0.114 (14.2)
BB	0.132 (21.0)	0.148 (20.0)	0.170 (19.9)	0.210 (19.3)
Cost	–0.161 (–22.2)	–0.158 (–21.5)	–0.167 (–21.2)	–0.196 (–19.9)
σ_ε				3.702 (7.5)
η			0.833 (44.9)	
SQ-Asc		0.290 (4.3)		–1.024 (3.1)
MRS_{AM}	0.42 (0.36,0.47)	0.50 (0.43,0.57)	0.57 (0.49,0.64)	0.58 (0.51,0.65)
MRS_{BB}	0.82 (0.6,1.01)	0.94 (0.8,1.03)	1.02 (0.8,1.14)	1.07 (0.8,1.26)
ln-L or ln-Sim$L$	–2,776	–2,766	–2,748	–2,719
AIC	5,558	5,540	5,504	5,450

Confidence intervals around marginal rates of substitution obtained with Krinsky and Robb (1986) method.

When the IIA property is not imposed (MXL-ε) the model fits the data best, and its estimates identify substantial positive correlation (0.9) amongst non SQ alternatives: the estimated total variance for non SQ utilities is 15.335, much larger than the Gumbel error variance of $\pi^2/6 \simeq 1.645$. This large variance is only in part surprising, as the public goods components in the attributes under valuation are of much more pervasive interest to the population of customers in study 2 than in study 1. Public goods are known to be subject to much larger variation in individual valuations than private goods. A negative Asc reveals that respondents are in fact more inclined to support change from the SQ. This attitude would be consistent with a perception of under-provision of the public goods under valuation. A similar finding is reported in Lehtonen *et al.* 2003.

In terms of the policy implications for amenity provision and quality of bathing waters the estimates imply the following. An average *WTP* per household between 0.42 and 0.57 pence for an increase in one unit in the number of areas with waste water discharges designed to allow recreational activities on rivers; and 0.81 and 1.07 pence per household for a one percent increase in the current government standards for bathing waters.

WTP estimates do not vary much in magnitude across models, with the notable exception of the obviously mis-specified MNL model, which provides lower point estimates. Considering the confidence intervals — obtained using the Krinsky and Robb method (1986) — a significant difference is observed only for the value estimates of AM.

6. Monte Carlo experiments

6.1 Monte Carlo Design

The analyses of the two data-sets lead to results with contrasting interpretations. The first set of results indicates that the three SQ specifications are statistically equivalent. The second highlights that differences across SQ specifications can be statistically significant, although—at this sample size—they are not so for the implied *WTP* estimates. This issue raises the question of evaluating the relative finite-sample performance of the three SQ specifications.

To explore such an issue we focussed on the effects of reciprocal mis-specifications in these models, and their sensitivity to sample size, by means of Monte Carlo (MC) experiments.

We ran a series of systematic experiments with GAUSS (routines are available from the authors) aimed at describing selected features of the finite sample properties of each of these specifications. The experiments were run using sample sizes (number of choices) that reflect those frequently encountered in the literature (N = 700, 1,400 and 2,900) so as to provide practitioners with some guidance about the expected efficiency gains achievable by increasing the sample size under different specifications and data generating processes.

Without loss of generality, we employ the data matrix of study 1 because it includes decreases of some valuable attribute levels. We use as data generating processes (DGPs) the set of estimates from this sample (Table 13.2), the only exception being the variance for the error component in MXL-ε. In this case the estimated σ_ε was not significant, so we choose to use a larger, yet realistic value (the estimated value from study 2, in Table 13.3). The steps involved are:
1) Compute the deterministic part of the utility for each alternative by using the maximum-likelihood (ML) or simulated-ML estimates reported in Table 13.2 and the original matrix of attribute levels $\mathbf{X}$.
2) Generate the unobserved stochastic component of the utility of each alternative by using pseudo-random draws (with seed) from the inverse cumulative distribution function suitable for each model.[13]
3) Derive an indicator of choice $\mathbf{y}^r$ from the alternative associated with the highest computed utility.

[13]Gumbel errors for models MNL, MNL-Asc and for the u_j of equation (3.8) the MXL-ε; GEV for the NL as for equation (3.5); and a re-scaled standard Normal for the ε_{c_1} and ε_{c_2} of equation (3.8) for MXL-ε.

4) Proceed to the estimation of the parameters of all models based on the simulated choice responses $\mathbf{y}^r$ and matrix of attribute levels $\mathbf{X}$, and save the relevant results of estimation (parameter estimates, t-values, log-likelihood at convergence etc.).
5) Repeat the previous steps for R=550 times.

Given the results in chapter 16 in this volume by G. Baiocchi, we report results of the MC experiment by using pseudo-random draws obtained with R and loaded into GAUSS. Overall the results were qualitatively similar to those previously obtained in GAUSS. The results presented here are from draws obtained from the free software R (those obtained in GAUSS are available from the authors upon request).

As a criterion to evaluate the performance of the various estimators we focus our attention on the expected difference between squared errors:

$$\Delta_{SE} = E\left((\hat{\gamma} - \gamma_0)^2 - (\tilde{\gamma} - \gamma_0)^2\right) \tag{6.1}$$

where $\hat{\gamma}$ is the estimator under mis-specification, $\tilde{\gamma}$ is the estimator correctly specified, and γ_0 is the true value from the DGP. The larger this value the worse is the consequence of mis-specification.

We prefer Δ_{SE} to the more frequently employed difference between mean squared errors:

$$E(\hat{\gamma} - \gamma_0)^2 - E(\tilde{\gamma} - \gamma_0)^2 \tag{6.2}$$

because of the lower variance associated with its estimator, which we estimate by:[14]

$$\bar{\Delta}_{SE} = \frac{1}{R}\sum_{r=1}^{R}\left((\hat{\gamma}_r - \gamma_0)^2 - (\tilde{\gamma}_r - \gamma_0)^2\right) \tag{6.3}$$

Since only relative values matter in the coefficient estimates in random utility models, we focus on the marginal rates of substitution ($\gamma =$ MRS), which are computed relative to the money coefficient. These measures—under certain conditions—can be interpreted as marginal *WTP* values, and hence are meaningful *per se*. Further, parameter estimates are asymptotically normally distributed, but MRS are non-linear functions and as such they do not have a well-defined sampling distribution.

[14]As Davidson and MacKinnon (1993) point out (page 740), the variance of equation (6.2) is: $R^{-1}V(\hat{\gamma}) + R^{-1}V(\tilde{\gamma}) - 2R^{-1}Cov(\hat{\gamma}, \tilde{\gamma})$ and for a positive covariance this variance is inferior to the variance associated with the difference of the mean squared errors. A positive covariance across estimates is very likely in our implementation because the estimates are obtained using the same pseudo-random draws.

Four types of concise measures are reported from the Monte Carlo experiments.[15]

First, we report the mean of the differences of the squared errors as from equation (6.2). Mis-specified models associated with large values of these are troublesome. Negative values indicate that the mis-specification is on average less biased than the correct estimator at that sample size, which can be explained by a compensating higher efficiency. To give a more readily interpretable measure of efficiency we also report the values of the inter-quartiles of these differences. The smaller these intervals the more efficient the mis-specification.

Secondly, we report the percent of cases in which the mis-specified estimator produces an estimate which is closer to the true value than the correct estimator. We report this in two forms, one for each MRS, $I(AF), I(RQ)$, and $I(OF)$ and one reporting the percent of cases in which this happens for *all three* attributes $I(AF, RQ, OF)$.

Thirdly, we report the mean of the relative absolute error:

$$\overline{RAE} = \tfrac{1}{R}\sum_{r}^{R}\left|\frac{\hat{\gamma}_r - \gamma_0}{\gamma_0}\right|. \tag{6.4}$$

This measure gives an idea of the relative magnitude of the bias of the estimate.

Finally, we report the fraction of MC experiments in which the estimated MRS is placed within a 5% interval around the true value, as a measure of efficiency computed as:

$$\Gamma_{0.05} = \tfrac{1}{R}\sum_{r}^{R} 1(\hat{\gamma}_r \in \gamma_0 \pm \gamma_0 \times 0.05). \tag{6.5}$$

Where $1(\cdot)$ is an indicator function. This count gives an idea of how clustered estimates are around the true values.

In addition, select points are illustrated using plots of the kernel smoothing of the obtained distributions of estimates, using the normal kernel with optimal bandwidth [4].

6.2 Monte Carlo Results

The results reported in tables 13.4-13.6 indicate that the values for $\bar{\Delta}_{SE}$[16] and their dispersion—as described by the size of the inter-quartile intervals in

[15] We omit to report the simulation performance of the AIC as a selection criterion for the correct specification. In brief, the simulation results showed that AIC was a stable indicator of performance and performed extremely well at all sample sizes and across all models.

[16] These values are scaled up by 1,000.

brackets —decrease as the sample size increases. Notice that in some cases—as evidenced by non-positive values of $\bar{\Delta}_{SE}$—the mis-specified model outperforms the true model in terms of the size of the expected squared-bias. This happens at all sample sizes and for all attributes when the true DGP is MNL-Asc and the specification is MXL-ε (Table 13.4). Under this DGP the specification MXL-ε seems to perform at least as well as the NL one, except at small sample sizes, and limited to $\Gamma_{0.05}$ and to individual $I(\cdot)$ values.

In terms of expected squared bias, when the DGP is NL the MXL-ε (Table 13.5) performs either as well (AF), or better than the correct specification at small sample sizes, but not at medium to high. Interestingly, at this sample size the MNL-Asc specification outperforms the true one for one attribute (OF). However, for this attribute the mis-specification MXL-ε gives more accurate estimates than the true specification 16% of the times, versus a 4 and 11% for the MNL and MNL-Asc, respectively. In terms of cases within the 5% interval around the true values, MXL-ε performs very similarly to the true specification at all sample sizes.

Notice, though, that the results in Table 13.6 show that when the true DGP is MXL-ε the mis-specifications *never* outperform the true specification in all the criteria, across all sample sizes. When, instead, MXL-ε was not the true DGP the mis-specifications never substantially outperform it. This is suggestive that, in the absence of a strong a-priori information on the true specification, the MXL-ε is preferable across the board.

In figure 13.1 we present a kernel plot of the distributions of the RAE for the *WTP* for Area Flooding when the true model is a nested logit, with N=2,900. From this figure it is evident how the real choice is between MNL, and the group MNL-Asc, NL and MXL-ε. Similar patterns emerge when the DGP is MNL-Asc, suggesting that these three models are effectively interchangeable. A stronger difference across specifications accounting for SQ emerges when the DGP is MXL-ε, as shown in figure 13.2. Here the true specification (dot-dashed line) shows a distribution of RAE values that outperforms the other two (dotted and continuous line) in that it is much more tightly concentrated on zero, while the MNL (dashed line) remains strongly biased.

7. Conclusions

Our empirical results from the analysis of the preferences of customers of Yorkshire Water are consistent with the fact that they are willing to pay for environmental improvements via an increase on their water bill. The estimated amount of *WTP* for quasi-public goods and pure public goods is plausible, and it is quite stable across the specifications used.

The models providing best statistical fit are found amongst those accounting for SQ effects, that was loosely defined as a systematic effect to choose the

Table 13.4. Summary statistics of Monte Carlo distributions of *WTP* estimates for DGP MNL-Asc.

	MNL			MNL-Asc			NL			MXL-ε		
Sample size	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900
$\bar{\Delta}_{SE}^{AF}$	3.33	2.49	2.8				0.12	0.15	0.16	–0.01	0	0
	(6.39)	(3.86)	(3.71)				(0.07)	(0.05)	(0.03)	(0.02)	(0.01)	(0)
$\bar{\Delta}_{SE}^{AF}$	5.39	26.09	26.52				0.96	0.87	1.02	–0.01	–0.04	–0.01
	(19.07)	(40.07)	(34.76)				(0.97)	(0.7)	(0.46)	(0.38)	(0.21)	(0.1)
$\bar{\Delta}_{SE}^{AF}$	–11.97	21.83	29.98				–1.56	–0.8	–0.4	–0.29	–0.02	–0.02
	(17.09)	(49.53)	(47.01)				(2.94)	(1.88)	(1.63)	(1.6)	(0.48)	(0.26)
$I(AF)$	9	4	0				56	53	48	55	53	51
$I(RQ)$	24	4	0				50	51	51	47	50	53
$I(OF)$	41	16	9				57	51	47	51	54	49
$I(AF, RQ, OF)$	4	1	0				18	22	21	20	22	21
$\overline{RAE}(AF)$	0.49	0.41	0.43	0.17	0. 10	0.07	0.17	0.1	0.08	0.17	0.1	0.06
$\overline{RAE}(RQ)$	0.19	0.32	0.32	0.13	0.08	0.05	0.13	0.08	0.06	0.13	0.08	0.05
$\overline{RAE}(OF)$	0.11	0.18	0.18	0.14	0.08	0.06	0.13	0.08	0.06	0.14	0.08	0.06
$\Gamma_{0.05}^{AF}$	0	0	0	17	34	46	19	32	45	18	34	45
$\Gamma_{0.05}^{RQ}$	9	0	0	25	38	55	25	39	54	24	37	55
$\Gamma_{0.05}^{OF}$	24	3	0	22	39	49	24	43	49	23	40	50

True absolute values of *WTP* are AF = 0.13, RQ = 0.52, OF = 0.97. The $\bar{\Delta}_{SE}$ values are multiplied by 1,000.

Table 13.5. Summary statistics of Monte Carlo distributions of *WTP* estimates for DGP NL.

	MNL			MNL-Asc			NL			MXL-ε		
Sample size	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900
$\bar{\Delta}_{SE}^{AF}$	3.38	2.57	2.85	0.02	0.01	0				0	0	0
	(6.23)	(4.02)	(3.83)	(0.12)	(0.04)	(0.03)				(0.07)	(0.45)	(0.05)
$\bar{\Delta}_{SE}^{RQ}$	5.14	27.19	27.1	0.07	0.05	0.02				–0.01	0.01	0.03
	(19.6)	(41.03)	(36.06)	(0.79)	(0.69)	(0.33)				(0.72)	(0.56)	(0.31)
$\bar{\Delta}_{SE}^{OF}$	–10.82	22.55	28.53	0.87	0.77	0.32				–0.44	0.07	0.07
	(16.25)	(50.21)	(48.16)	(5.58)	(2.98)	(1.68)				(3.3)	(1.9)	(1.11)
$I(AF)$	1	2	0	43	43	43				50	47	45
$I(RQ)$	24	1	0	47	51	47				48	47	46
$I(OF)$	42	15	5	43	41	45				50	47	49
$I(AF, RQ, OF)$	4	0	0	11	11	12				16	18	18
$\overline{RAE}(AF)$	0.48	0.41	0.43	0.16	0.09	0.07	0.16	0.09	0.07	0.16	0.1	0.07
$\overline{RAE}(RQ)$	0.18	0.32	0.31	0.13	0.07	0.05	0.13	0.07	0.05	0.13	0.07	0.05
$\overline{RAE}(OF)$	0.12	0.18	0.18	0.14	0.09	0.06	0.14	0.08	0.06	0.14	0.08	0.06
$\Gamma_{0.05}^{AF}$	0	0	0	19	37	41	20	37	41	20	36	41
$\Gamma_{0.05}^{RQ}$	10	0	0	25	45	56	26	44	56	25	43	56
$\Gamma_{0.05}^{OF}$	23	4	0	25	35	50	25	36	51	25	35	51

True absolute values of *WTP* are AF = 0.13, RQ = 0.52, OF = 0.97. The $\bar{\Delta}_{SE}$ values are multiplied by 1,000.

Table 13.6. Summary statistics of Monte Carlo distributions of *WTP* estimates for DGP MXL-ε.

	MNL			MNL-Asc			NL			MXL-ε		
Sample size	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900
$\bar{\Delta}_{SE}^{AF}$	8.12	6.21	6.55	0.31	0.14	0.07	0.28	0.13	0.06			
	(13.98)	(8.87)	(8.28)	(1.22)	(0.55)	(0.31)	(1.14)	(0.53)	(0.27)			
$\bar{\Delta}_{SE}^{AF}$	13.56	38.95	40.44	2.89	3.15	1.21	6.31	2.45	0.88			
	(38.55)	(60.17)	(54.96)	(14.18)	(13.86)	(5.45)	(17.9)	(9.72)	(3.96)			
$\bar{\Delta}_{SE}^{AF}$	–13.8	28.16	30.61	13.68	8.57	4.96	17.86	13.59	10.32			
	(19.95)	(65.16)	(55.3)	(53.66)	(33.41)	(19.93)	(79.52)	(48.89)	(33.35)			
$I(AF)$	2	0	0	37	31	32	35	35	36			
$I(RQ)$	22	2	0	39	38	41	31	38	39			
$I(OF)$	58	15	8	42	40	38	42	37	31			
$I(AF, RQ, OF)$	2	0	0	8	6	6	5	9	8			
$\overline{RAE}(AF)$	0.73	0.64	0.65	0.21	0.13	0.1	0.21	0.13	0.09	0.17	0.17	0.08
$\overline{RAE}(RQ)$	0.26	0.39	0.39	0.17	0.13	0.09	0.19	0.13	0.08	0.14	0.1	0.07
$\overline{RAE}(OF)$	0.14	0.2	0.19	0.19	0.12	0.09	0.19	0.13	0.11	0.16	0.09	0.07
$\Gamma_{0.05}^{AF}$	0	0	0	16	25	32	14	27	34	19	30	38
$\Gamma_{0.05}^{RQ}$	6	0	0	18	24	39	18	26	43	23	31	45
$\Gamma_{0.05}^{OF}$	19	4	0	16	25	38	18	25	30	16	33	46

True absolute values of *WTP* are AF = 0.13, RQ = 0.52, OF = 0.97. The $\bar{\Delta}_{SE}$ values are multiplied by 1,000.

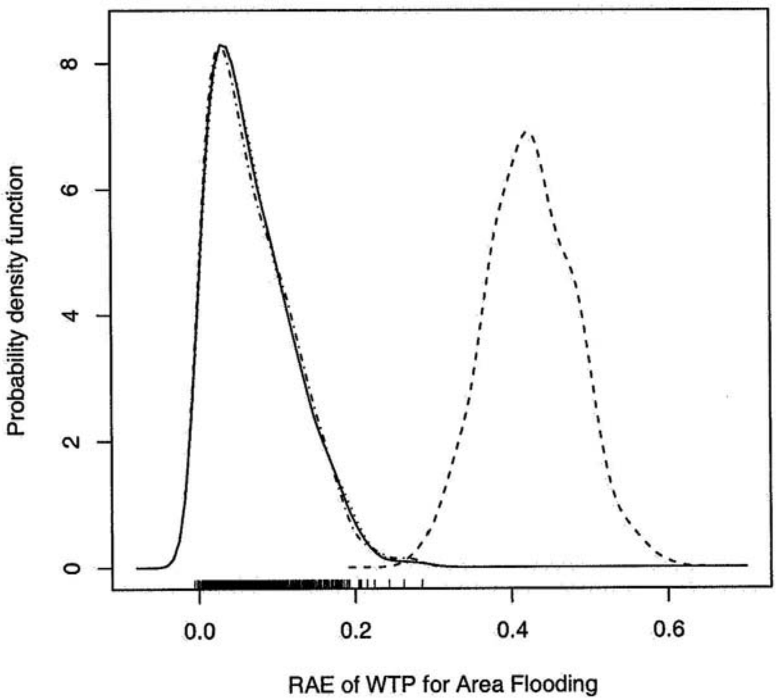

Figure 13.1. Plot of kernel-smoothed distribution of the Relative Absolute Error of *WTP* estimates for Area Flooding. True data generating process (Continuous line NL): MXL-ε dashed-dotted line; Dashed line MNL; Dotted line MNL-Asc.

SQ or the alternatives different from the SQ beyond what can be explained on the basis of the attributes values alone. We found that in our samples the conditional logit model, that ignores any source of SQ effect, produces the lowest estimates of benefits from provision of externalities. While from the societal viewpoint such a conservative estimate would guide investments in a cautious way, it would still represent a sub-optimal resource allocation, as many potentially beneficial proposals would fail the Pareto efficiency test by providing too low a benefit estimate.

Following other authors (Haaijer, 1999; Kontoleon and Yabe 2003), we have argued that there are very good reasons for investigating the existence of SQ effects in the application of choice-experiments, and that these reasons might be particularly compelling in non-market valuation of environmental goods.

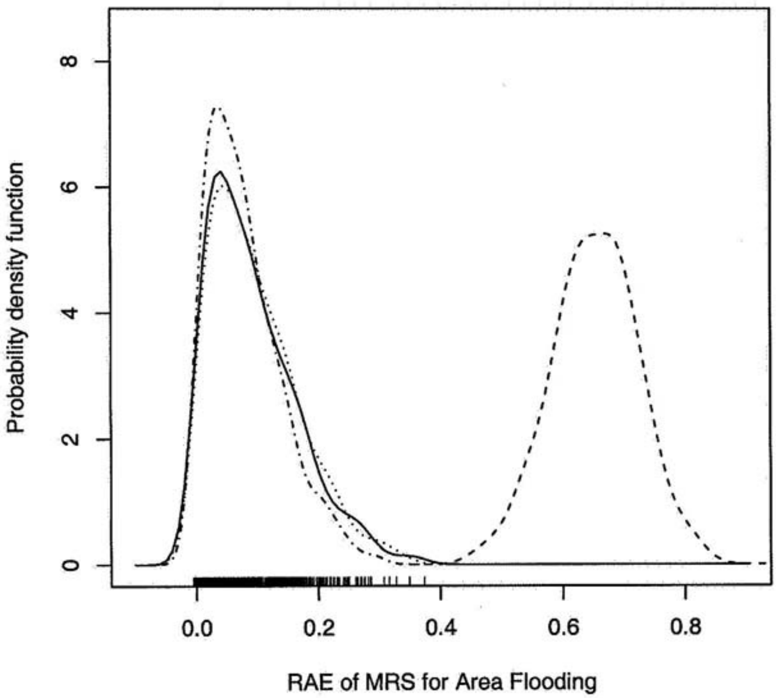

Figure 13.2. Plot of kernel-smoothed distribution of the Relative Absolute Error of *WTP* estimates for Area Flooding. True data generating process: MXL-ε (dashed-dotted line). Continuous line NL; Dashed line MNL; Dotted line MNL-Asc.

We examined three specifications that can be used to account for these effects: the conventional logit model with alternative-specific constant, the nested logit model and and the less conventional mixed logit with error components and alternative specific constant.

Secondly, we reported how we observed different forms of statistical evidence of SQ effects in two separate studies on preferences for water management attributes, which include important public goods, such as number of areas protected by flooding and number of households protected from odour and flies. While in a study we observe that all three specifications accounting for SQ afford similar statistical performance and *WTP* estimates, in the other application we observe that the mixed logit with error component and alternative-specific constant statistically dominates the nested logit and MNL-Asc, but this dominance does not implies statistically different estimates.

Finally, we investigated the effects of mis-specification using in turn the three SQ data generating processes by means of Monte Carlo experiments over a plausible range of sample sizes. The results of the experiments suggest a number of points.

First, when SQ effects are a concern, the use of simple conditional logit specifications may produce strongly biased estimates for the taste parameters. These will also produce biased welfare measures.

Secondly, when the true DGP is mis-specified, the MXL-ε specification generally provides a good performance in our Monte Carlo experiments. Such performance is not matched neither by the NL model nor by the MNL-Asc model when the true DGP is MXL-ε.

In conclusion, our empirical results confirm the existence of a systematic effect of the status-quo alternative on choice selection. This was previously discussed and evidenced in general terms by Samuelson and Zeckhauser (1988) and Hartman *et al.* (1991). Such effect was examined more specifically in the context of choice-experiment in market research by Haaijer (1999) and Haaijer *et al.* (2001) and addressed in environmental economics by Hanley and Wright (2003), and Li *et al.* (2004) by means of nested logit models.

We find that a less usual specification, namely the MXL-ε consistently achieves better results than MNL with an alternative-specific constant for the SQ and NL specifications. The MXL-ε model is parsimonious, yet, it captures SQ effects in both the systematic component of preference via alternative-specific constant, and the unobserved heterogeneity associated with hypothetical changes described by unfamiliar attribute levels. It also breaks away from the restrictive independence of irrelevant alternatives.

Of course the usual caveats pertaining to Monte Carlo results apply here. Namely, these results might be not very general and perhaps they are due to the particular data employed in this study. Nevertheless we find quite plausible that a specification that accommodates status-quo effects simultaneously in both the stochastic and deterministic component of utility outperforms specifications that only address one at the time.

Further research should investigate how general these preliminary results are, and how status-quo effects can be related to the various features of the experimental design, investigating — for example — the relationship between choice-complexity and degree of familiarity with attributes levels defining the status-quo vis-à-vis the proposed changes.