Edited by
Duangkamon Chotikapanich · Alicia N. Rambaldi ·
Nicholas Rohde

Advances in Economic Measurement

A Volume in Honour of D. S. Prasada Rao

MOREMEDIA

palgrave macmillan

Advances in Economic Measurement

Duangkamon Chotikapanich ·
Alicia N. Rambaldi · Nicholas Rohde
Editors

Advances in Economic Measurement

A Volume in Honour of D. S. Prasada Rao

palgrave
macmillan

Editors
Duangkamon Chotikapanich
Monash University
Melbourne, VIC, Australia

Alicia N. Rambaldi
The University of Queensland
St Lucia, QLD, Australia

Nicholas Rohde
Griffiths University
Southport, QLD, Australia

ISBN 978-981-19-2022-6 ISBN 978-981-19-2023-3 (eBook)
https://doi.org/10.1007/978-981-19-2023-3

This Palgrave Macmillan imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Economic Measurement, the theme for the set of papers published in this Volume, is of critical importance in a world where policymakers, businesses and individuals are increasingly reliant on evidence-based decision-making. Economic statistics such as the consumer price index (CPI), gross domestic product (GDP), per capita income and growth rates are used by the general public, as well as analysts and policymakers at the central banks setting and evaluating monetary policy. Theoretical and applied econometric techniques developed over the last century underscore the role of modern economic measurement for policymaking in complex settings, and in evaluating the effectiveness of policy interventions for improving health and educational outcomes in developing countries.

Determining the scope and identifying strands of economic measurement to be covered by this edited Volume is a challenging task. The editors have shown considerable wisdom and finesse in selecting three strands of economic measurement which are intricately connected to measurement of economic welfare. The size of the economy and its distribution are the core determinants of economic welfare, and the long-run sustainability of economic growth is inevitably determined by productivity growth performance of the economy. Measurement of efficiency and productivity and recent advances in this important direction form the first theme of this volume. It has long been acknowledged that the size of the economy has a strong bearing on the economic welfare as it reflects the command over goods and services, but it is not an adequate measure of national welfare. Consequently, the second theme for this volume focuses on inequality in the distribution of income, health as well as the more fundamental problem of inequality of opportunity. The third focuses on measures of the size of an economy, including the gross domestic product, and comparisons over time and space. Through the selection of these three welfare-related topics in economic measurement, the editors have also managed to align the contributions included in this volume with the strands of research that have kept me occupied through my long academic career spanning nearly fifty years.

The authors of the chapters in this Volume are leading researchers in their respective fields of inquiry. These chapters provide an overview of the literature and the world's best practice as well as the considered wisdom of the contributors. I sincerely hope that the readers find these chapters just as fascinating, informative and useful as I have.

"*The Journey, Not the destination matters…*" (T.S. Eliot). How true! It is certainly the case with my own academic journey which started at the Indian Statistical Institute, Calcutta and ended at the University of Queensland, Brisbane, Australia with a long stopover at the University of New England, Armidale, Australia. I am indeed fortunate for having had the opportunity to meet and work with outstanding researchers and to supervise some brilliant young scholars. Along the way, I was able to establish fruitful research collaborations but more importantly make lasting friendships all over the world. I am indeed grateful to Duangkamon Chotikapanich and Nicholas Rohde who are among the best of my Ph.D. students and a long-time colleague and friend Alicia Rambaldi for undertaking this arduous and labour-intensive task of bringing this Volume together. I am deeply indebted to the contributors of this volume for the time and effort they have put into crafting these masterful chapters. I wish to conclude by assuring all my friends that I am officially retired, but not quite retired from the academic pursuits I so dearly love!

D. S. Prasada Rao
Emeritus Professor
The University of Queensland
St Lucia, QLD, Australia

Preface

The purpose of this book is to honour D.S. Prasada Rao and his many outstanding contributions to economic measurement, including: index number methods for international comparisons of prices, real incomes, output and productivity; stochastic approaches to index numbers; purchasing power parities for the measurement of regional and global inequality and poverty; and the measurement of income and economic insecurity.

Prasada obtained his B.A. in 1964 and M.A. in 1966 from Andhra University, and his Ph.D. in 1973 from the Indian Statistical Institute. His first position was as Lecturer in 1974 at the Department of Econometrics at the University of New England in Australia, where he rose to Professor in 1997, and was founding Director of the Centre for Efficiency and Productivity Analysis (CEPA) in the year 2000. From there, he went to the Department (now School) of Economics at the University of Queensland as Professor of Econometrics and Director of CEPA (2003–2008). He was ARC Professorial Fellow between 2009 and 2013, and continued as Professor of Econometrics at the School of Economics until 2018. He has been Emeritus Professor since 2018. Prasada has also held visiting positions at various universities around the world. He is Fellow of the Academy of Social Sciences in Australia, and Fellow of the Society for Economic Measurement.

During his long and distinguished career, Prasada has published over 95 research papers, 26 book chapters and 12 books. His lasting contributions have influenced how the world measures and tracks inequality, poverty and productivity. He has undertaken several global-scale empirical studies involving a large number of countries for the FAO, ILO and the World Bank. His work on inter-country comparisons of agricultural output and productivity for the FAO has been influential in the compilation of the FAO Production Index Numbers, and it has provided the basis for considerable research on global agricultural output and productivity. He played a very active role in the 2005, 2011 and 2017 International Comparison Programs (ICP) at the World Bank, as well as at the Asian Development Bank, the regional coordinating agency for ICP in Asia and the Pacific. In addition, he has overseen a major research project on the Measurement of Purchasing Power Parities (PPPs) for Global Poverty Measurement for the Asia-Pacific Region. He has been appointed as Member of the Technical Advisory Group for the ICP at the World Bank since 2003. Through collaborations with Chotikapanich, Griffiths and Hajargasht, he developed rigorous methodological techniques for modelling and estimating national income distributions from limited data, tools that can be used for assessing poverty and inequality at national, regional and global levels. His collaboration with Rambaldi, Doran, Hajargasht and Balk have led to the development of methodologies to compute standard errors for the ICP PPPs, time-space consistent panels of PPPs with standard errors, real incomes at current and constant prices, and global and regional measures of growth and inflation. Over his long career, he supervised to completion 19 Ph.D. students, among them Chotikapanich and Rohde.

This book is a collection of papers written by well-known and influential researchers in the fields to which Prasada has made significant contributions. His standing in these fields has enabled us to attract world-leading frontier researchers to contribute to the volume. The papers are grouped into three parts, each of which relates to an area in which Prasada has made major contributions. Papers in Part I are concerned with various aspects of efficiency and productivity measurement. Part II contains papers on income distribution, welfare inequality and insecurity. The papers in Part III cover index numbers and international comparisons of prices and real expenditures. The contributed papers review the existing methods and applications as well as some recent developments.

We would like to thank Palgrave Macmillan and Springer Nature for giving us the opportunity to honour Prasada in this way. To those who used their valuable time refereeing the papers, we also say thank you.

Melbourne, Australia	Duangkamon Chotikapanich
St Lucia, Australia	Alicia N. Rambaldi
Southport, Australia	Nicholas Rohde

Contents

List of Contributors

Apouey Bénédicte Paris School of Economics—CNRS, Paris, France

Balk Bert M. Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands

Chotikapanich Duangkamon Monash University, Melbourne, VIC, Australia

D'Ambrosio Conchita University of Luxembourg, Esch-sur-Alzette, Luxembourg

Diewert W. Erwin Vancouver School of Economics, University of British Columbia, Vancouver, BC, Canada;
School of Economics, University of New South Wales, Sydney, NSW, Australia

Griffiths William University of Melbourne, Melbourne, VIC, Australia

Hajargasht Gholamreza School of Economics, University of Queensland, Brisbane, QLD, Australia

Hill Robert J. Department of Economics, University of Graz, Graz, Austria

Inklaar Robert Groningen Growth and Development Centre, University of Groningen, Groningen, The Netherlands

Lovell C. A. K. Centre for Efficiency and Productivity Analysis, School of Economics, University of Queensland, St Lucia, QLD, Australia

Marapin Ryan Groningen Growth and Development Centre, University of Groningen, Groningen, The Netherlands

Nguyen Bao Hoang School of Economics, The University of Queensland, Brisbane, QLD, Australia

Palmisano Flaviana Sapienza University of Rome, Rome, Italy

Peragine Vito University of Bari, Bari, Italy

Peyrache Antonio Centre for Efficiency and Productivity Analysis (CEPA), School of Economics, University of Queensland, St Lucia, QLD, Australia

Rambaldi Alicia N. School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, St Lucia, QLD, Australia

Rohde Nicholas Department of Accounting, Finance and Economics, Griffith Business School, Southport, QLD, Australia

Schreyer Paul OECD, Paris, France

Sickles Robin C. Economics Department, Rice University, Houston, TX, USA

Silber Jacques Bar-Ilan University, Ramat Gan, Israel

Silva Maria C. A. CEGE—Católica Porto Business School, Porto, Portugal

Timmer Marcel Groningen Growth and Development Centre, University of Groningen, Groningen, The Netherlands;
Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands

Watson Barry University of New Brunswick, Fredericton, NB, Canada

Woltjer Pieter Groningen Growth and Development Centre, University of Groningen, Groningen, The Netherlands

Zelenyuk Valentin School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, Brisbane, QLD, Australia

List of Figures

List of Tables

List of Boxes

PART I

Productivity Measurement

CHAPTER 1

Productivity Measurement: Past, Present, and Future

C. A. K. Lovell

When it is obvious the goals cannot be reached, don't adjust the goals, adjust the action steps.

Confucius (551BC–479BC)

All things will be produced in superior quantity and quality, and with greater ease, when each man works at a single occupation, in accordance with his natural gifts, and at the right moment, without meddling with anything else.

Plato (428BC–348BC)

What we measure affects what we do…if we don't measure something, it becomes neglected.

Stiglitz et al. (2018)

C. A. K. Lovell (✉)
Centre for Efficiency and Productivity Analysis, School of Economics, University of Queensland, St Lucia, QLD, Australia
e-mail: k.lovell@uq.edu.au

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_1

Introduction

Confucius, Plato, and Stiglitz and colleagues all made very different, and equally astute, observations about productivity, and long before Stiglitz and colleagues, the distinguished management gurus W. Edwards Deming and Peter Drucker were both alleged to have claimed along similar lines that that we can't manage what we don't measure. In this survey, I pursue these observations, and more. The survey considers the goals or objectives of economic agents, the action steps and methods they follow, the measured and unmeasured outcomes of their production activities, and how to measure the former and incorporate the latter, each as it pertains to the productivity of economic agents ranging from businesses to national economies. The middle of the year 2021 is a good time to think about productivity, its drivers and its impacts, in light of the unprecedented simultaneous challenges presented by the pandemic depression and climate change.

This chapter is structured as follows. In section "From the Past to the Present", I provide an overview of productivity in the past, in some cases of economic features closely related to productivity, from the distant past to the recent past, covering the first era in the title of this chapter. In section "The Distant Past: Observation from Antiquity to Adam Smith, Alfred Marshall and A. C. Pigou", I conduct a quick trip through the distant past, beginning with Antiquity and focused primarily on observation, including those of Smith, Marshall, and Pigou. In section "A Mere Century Ago: Accumulating Methods and Evidence Amidst Emerging Social Concerns", I examine four significant achievements that began a mere century ago. The first involves progress from ancient observation to the development of index numbers suitable for productivity measurement. The second surveys applications of index numbers to productivity measurement, beginning with labour productivity and continuing with total productivity (I prefer Kendrick's qualifier "total" to the more popular "total factor" or "multifactor"). The third involves a movement, born between World Wars I and II and maturing in the aftermath of the Great Depression, that proposed an expansion of the focus of measurement, from narrow economic productivity to holistic social economic performance, beginning with the ostensibly adverse social impacts of the introduction of new labour-saving technology. The final achievement belongs to the short-lived European Productivity Agency and its house journal *Productivity Measurement Review*, which for a decade published a stream of articles documenting productivity trends and productivity dispersion, primarily among plants or firms.

In section "Converging to the Present: Analytical Foundations and Drivers", I examine four significant achievements in productivity measurement that have occurred in the present, in which I include the recent past that began in the middle of the last century and continues to the present day. In section "Analytical Foundations of Productivity Measurement", I recount the development of a suite of theory-based analytical foundations of productivity measurement. These analytical foundations support alternatives to index numbers to measure productivity. In section "Drivers of Productivity Change", I survey a development that goes beyond measurement to a search for the drivers of, and impediments to, productivity growth. I assign drivers and impediments to five categories: quality, technology, organisation, institutions, and geography. In section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies", I explore a rich literature devoted to an analysis of productivity dispersion, from two occasionally intersecting perspectives, one focusing on resource misallocation and reallocation, the other focusing on productivity gaps, distance to a best practice productivity frontier, and the identification of zombie firms and their zombie jobs. Both perspectives occur primarily though not necessarily among firms, with adverse impacts on aggregate productivity. In section "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth", I explore a revival of the call for expanding the scope of productivity measurement to incorporate social impacts, the revival going by a variety of names including inclusive green growth.

I introduce section "The Future: Confronting Two Challenges of Transcendent Significance" by observing that at this point we have amassed data, we have analytical foundations and empirical techniques, and we have experience gained from applying techniques to data in a rich variety of environments and circumstances. I then speculate on how well the developments accumulated to date have prepared us to confront the future, by reporting the latest information and conjecture on two challenges of transcendental significance. Much of the relevant literature is new, and this section is replete with references to working papers. In section "Productivity and the Pandemic Depression", I discuss the COVID-19 pandemic, the deep depression it has spawned, and its relation to productivity. In section "Productivity and Climate Change", I discuss the growing awareness of climate change and its relation to productivity.

In section "Linkages Between the Two Challenges", I discuss the linkages between the two challenges, and the policy options for a green recovery. The literature on these two challenges and their intersection has accumulated rapidly and is accelerating, and so I arbitrarily conclude this subject in the middle of 2021. It will be educational to revisit section "The Future: Confronting Two Challenges of Transcendent Significance"" when the sun finally has set on the pandemic depression. Lovell (2021) provides an expanded version of section "The Future: Confronting Two Challenges of Transcendent Significance".

Finally, in section "Conclusions", I summarise the survey, take stock of some important omissions, and hazard a look ahead.

A central message of this unconventional survey is that the purview of productivity analysis and measurement has widened greatly through its development, and continues to widen, from its original mainstream focus on some aggregate measure of national income per capita or per worker. Productivity analysis has developed in four general directions, each with a variety of interesting offshoots. First, it has sought, with limited empirical success to date, to incorporate a more holistic sense of what an economy generates with its limited resources, and what these resources include. At the aggregate level, national income has been augmented to include various indicators of economic well-being, or social economic progress, or inclusive green growth, or even Gross National Happiness. Some of these developments appear in sections "Social Concerns" and "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth". At the individual business level, it is not what is being produced, but rather the distribution of the income generated, that has been broadened. An expansion of the distribution of the fruits of productivity growth from shareholders to myriad stakeholders has been proposed, as businesses have been endowed with new holistic objectives of corporate social responsibility (CSR) or environment, social, and governance (ESG). I regret not covering these micro developments in this survey, because they have both financial and productivity implications, but the literature is huge, growing rapidly in the wake of recent initiatives of the US Business Roundtable (https://www.businessroundtable.org/policy-perspectives/corporate-governance) and the British Academy (https://www.thebritishacademy.ac.uk/programmes/future-of-the-corporation/), and easy to find. Second, and perhaps more feasibly from an empirical perspective, productivity analysis has harnessed economic theory to develop an analytical approach to productivity measurement that more accurately reflects the variety of objectives and constraints actually faced by producers. Objectives range far afield from the textbook profit maximisation goal, and many constraints are imposed externally, by the institutional environment in which business operates. Constraints vary across jurisdictions, and they influence aggregate productivity in two ways, by reducing the productivity of individual producers, and by creating productivity dispersion among producers. Some of these analytical developments appear in section "Analytical Foundations of Productivity Measurement", and empirical applications appear throughout the survey, most extensively in sections "Drivers of Productivity Change" and "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies". Third, analytical techniques have been applied to increasingly detailed databases to measure productivity, the shape and moments of its distribution, and its convergence or divergence through time, at both individual firm and aggregate levels. Empirical findings of large and often growing dispersion have spurred interest in the role of public policies that might reduce dispersion and increase aggregate productivity. These developments also appear in sections "Drivers of Productivity Change" and "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies". Fourth, productivity analysis has been applied to an investigation into the causes and consequences of economic depressions, especially those brought on by pandemics, and to an analysis of the drivers and impacts of climate change. There exists no more relevant example than the situation we find ourselves in at the beginning of the third decade of the twenty-first century, which I examine in section "The Future: Confronting Two Challenges of Transcendent Significance".

I conclude this Introduction with some guidance for the reader. What follows is not a conventional survey of productivity measurement, assuming such a thing exists, but rather an overview of where the literature has been and where it is likely to be headed, both guided and constrained by my own research interests. I intend it to provide a readers' guide to a somewhat idiosyncratic literature in which productivity has played, or should have played, or may yet play, a key role, and even in some situations when productivity has played little or no role. The idiosyncrasy of the survey is apparent, for example, where I pay more than passing attention to the role of business management and societal institutions in influencing productivity, and where I consider seriously the advantages and drawbacks of expanding the scope of productivity analysis beyond the market economy. It is also reflected in my reliance on unconventional sources in addition to academic books and journal articles for insights and information. Conventional, although dated to varying degrees, surveys appear in Hulten (2001), Syverson (2011), and Grifell-Tatjé et al. (2018a).

From the Past to the Present

In this section. I cover a lot of chronological ground, from antiquity through the Middle Ages to the dawn of the twentieth century in section "The Distant Past: Observation from Antiquity to Adam Smith, Alfred Marshall and A. C. Pigou". In section "A Mere Century Ago: Accumulating Methods and Evidence Amidst Emerging Social Concerns", I discuss four significant twentieth-century developments, the construction of index numbers with which to measure productivity change, the use of index numbers to gather evidence, the birth of a movement to broaden the scope of productivity measurement, and the brief but influential life of the European Productivity Agency.

The Distant Past: Observation from Antiquity to Adam Smith, Alfred Marshall and A. C. Pigou

There was little or no concept of measurement, of resources, production, or productivity, in the distant past, but there did exist a practice of observation that guided subsequent research and, at least in the case of Alfred Marshall, produced testable hypotheses. I briefly consider modern research into Antiquity and Maddison's Merchant Capitalist and Capitalist epochs in the first three subsections, with an objective not of offering a complete account, but of providing a sense of what has been possible with such limited resources. I summarise some relevant writings of three giants from the late eighteenth century through the early twentieth century in the final subsection.

As a preview of coming attractions, Maddison (2006) reported a 13-fold increase in per capita income over the past millennium. He attributed the growth in this indicator of economic performance to three interactive processes: conquest or settlement of relatively empty areas, international trade and capital movements, and technological and institutional innovation. These three processes appear frequently in this brief survey of the distant past.

Observation in Antiquity

The absence of measurement at the time created a paucity of data that has severely limited the ability of modern writers to analyse the performance of ancient economies. To cite just one example, White (1956), referring to antiquity in general and Roman agriculture in particular, lamented the absence of detailed statistical information on which to base an accurate assessment of economic performance, including such essential information as quantities of crops, labour input to each, and average yields per acre of each. The paucity of data is one of four themes permeating modern research into antiquity; consequently, the second theme is the nearly complete absence of the word "productivity". The third theme is the recurrence of culture, institutions, and technology as drivers of and impediments to ancient economic activity. The fourth theme, one that does not require detailed information, is the role played by location; inland agriculture and manufacturing were hampered by inadequate facilities for land transport, while transport by river and sea was neither costly nor inefficient. The third and fourth themes reappear in section "Drivers of Productivity Change".

The data constraint notwithstanding, keen observers have learned much and written widely about ancient economies. Adam Smith (1776; Book II, Chapter V) was an early observer of antiquity, writing of the opulence and industriousness of the ancient Greek, Egyptian, Chinese, and Indostan empires, "…the wealthiest, according to all accounts, that ever were in the world…", attributing the wealth of the latter three to their superiority in agriculture and manufacturing, although not in foreign trade. He cited Montesquieu, who wrote that the Egyptians had a superstitious antipathy to the sea. In a recent study of ancient Greece, Tridimas (2019) attributed its growth and prosperity to its institutions, primarily citizenship and the enforcement of property rights, its culture, including a positive approach to work, competition, and the accumulation of wealth, and its location, which gave it access to external trade. He attributed its eventual decline to its many small city-states, an organisational structure that prevented the exploitation of scale economies that would have fostered continued growth. These city-states were often at war, requiring a reallocation of resources that sapped their growth potential. In his study of the ancient world, Greene (2000) recounted steady economic growth in the Greek states and the Roman empire over an extended period, which he attributed to the existence of legal, administrative, and financial institutions that enhanced overall economic activity, and the extensive exploitation of slave labour and the spread of new technology, particularly in agriculture and mining. Norberg (2020) emphasised the significance of Rome's coastal location, writing that not only roads led to Rome, but shipping lanes as well, and attributed Roman prosperity to its location and its openness to trade, people, and ideas. Taken together, these studies illustrate the roles played by the three fundamental sources of economic growth identified by Acemoglu et al. (2005) as institutions, culture, and geography.

C. Clark (1940 [1951, Excursus]) actually managed to find some quantitative evidence, although not of productivity. He quoted extensively from Rostovtzeff (1926), who wrote of epochs of "high economic development" and "complex economic life" achieved by ancient Egypt, Babylonia, Rome, and Athens based on their large internal markets, advanced production techniques, pure and applied science, and slave labour, impeded only by constant warfare and augmented in Athens by growing external trade. Clark then amassed a wide array of empirical evidence in support of Rostovtzeff's evaluation, consisting mostly of price and wage data, including data enumerating the purchasing power of the Greek drachma beginning in 400–375 BC, and the Greek real standard of living at about the same time.

Summarising, even now we have no information on productivity levels or changes in the ancient world. However, we do have a wealth of observation, and very limited quantitative information, on prosperity and some of its sources, all suggestive of a relatively high level of economic performance. It is not difficult to imagine that some of that high performance reflects high productivity, and that variation in performance across space and through time reflects productivity growth and decline.

Evidence, 1500–1820 (Maddison's Merchant Capitalist Epoch)

Maddison (2006) gathered a massive amount of information on the world economy, from year 1 AD through 1998. For the endpoints of his merchant capitalist epoch 1500–1820, he reported levels and average annual rates of growth of GDP per capita, in 1990 international dollars using Geary-Khamis multilateral PPPs, for 20 countries, eight regions, and the world. Depending on how closely employment tracked population during this period, GDP per capita provides a workable approximation to a measure of labour productivity. Regional GDP per capita levels varied widely in 1500, and national levels even more so. Average annual rates of growth of GDP per capita also varied widely across regions and more widely across nations. Maddison also discussed in some detail three proximate causes of economic growth and its variability. Conquest and settlement brought new fertile land and biological resources and a potential transfer of population, crops, and livestock. International trade and capital movements expanded domestic markets that had limited the division of labour and allowed a transfer of technology. Trade also enhanced the discovery and dissemination of new technology, particularly in agriculture and maritime navigation.

Other studies of the period exist, including those of Allen (2000, 2001), who studied "the great divergence" in European real wages between 1500 and 1750 and dispersion in European agricultural labour productivity from 1300 to 1800, but the message is consistent. It is one of wide dispersion, through time (and often in the wrong direction) and across countries. This productivity dispersion chronicled by Maddison and Allen remains with us today; see section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies".

Evidence, 1820–1913 (Maddison's Capitalist Epoch)

Maddison (2006) reported levels and average annual growth rates of GDP per capita, also in 1990 international dollars using Geary-Khamis multilateral PPPs, during 1820–1870, 1870–1913, and three subsequent time periods in his capitalist epoch for countries and regions (e.g., Western Europe, Western offshoots, etc.). World per capita GDP growth accelerated relative to the Merchant Capitalist Epoch, from 0.05% pa in 1500–1820 to 0.53% pa in 1820–1870 and 1.30% pa in 1870–1913. The Western Offshoots (US, Canada, Australia, and New Zealand) led the way, followed by Western Europe in 1820–1870 and Latin America in 1870–1913. Inter-country and inter-region dispersion continued, characterised by divergence rather than convergence. Maddison (2005; Table 7b) was able to report average annual growth rates of total productivity for just three countries, the UK, the US, and Japan, for the same time periods, with similar results.

Earlier scholars lacked data and techniques available to Maddison, but nonetheless produced considerable evidence on productivity around the world. C. Clark (1940 [1951]) reported that real national income per person in work and per hour worked nearly doubled and more than doubled, respectively, in the US from 1800 through 1913. These two measures of labour productivity nearly tripled and more than doubled in France during the same period. Roughly similar findings for other countries generally refer to shorter time periods within the Capitalist Epoch. Fourastié (1951) amassed a different sort of evidence, much more detailed and more narrowly focused on France from 1830. In addition to real national income per capita, which increased nearly fourfold through 1900, he reported hourly and daily wage rates of unskilled labour and prices of a wide range of consumption goods, from which he constructed price-based productivity indices; see section "Methods" for a brief treatment of price-based productivity indices. With an eye toward what he called the level of living, he also reported trends in the number of doctors and dentists, consumption of various commodities, and several modes of transport services available.

Observations of Adam Smith, Alfred Marshall, and A. C. Pigou

Adam Smith (1776 [1937]) scattered apt remarks and observations throughout *The Wealth of Nations*. He devoted Book II, Chapter III to productive and unproductive labour and the accumulation of capital, where he got to the heart of labour productivity and two of its determinants. I am not the first to quote Smith on the matter Spengler (1959; 405) and Kendrick (1961; 3) predate me by a wide margin with identical quotations], but the following phrase differs sufficiently from previous quotations to justify its place here:

> The annual produce of the land and labour of any nation can be increased in its value by no other means, but by increasing either the number of its productive labourers, or the productive powers of those labourers…The productive powers of the same number of labourers cannot be increased, but in consequence either of some addition and improvement to those machines and instruments which facilitate and abridge labour; or of a more proper division and distribution of employment.

Smith then wrote of "…perversion of the annual produce from maintaining productive to maintain unproductive hands…" and of "…absolute waste and destruction of stock…", both of which retarded capital accumulation; in today's parlance, they created productivity gaps, a topic I survey in section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies". However, Smith's fundamental insight was that the productivity of labour can be increased in three ways, by employing additional complementary capital, by improvements in technology, or, famously, by the division and reallocation of labour. This insight encourages me to skip back to Smith's Book I, Chapters I–III. There he wrote in some detail and depth of the importance of the division of labour ("[t]he greatest improvement in the productive powers of labour…seem to have been the effects of the division of labour."), its origins in improved dexterity, time savings and innovation, and its limitation by the extent of the market, which can be expanded through improvements in transportation and infrastructure (he emphasised the significance of water-carriage) and lowered trade barriers. In today's literature, the division of *labour* is replaced by the reassignment of *tasks*, with the same importance, as I explore in section "Social Concerns".

Leaping ahead a century, Alfred Marshall (1890 [1961]), like Smith before him, wrote of the economic strengths and weaknesses of the ancient Greek and Roman civilisations. However, his relevant contributions came in Book IV, in which he wrote of the agents of production, which he counted as two (nature and man), three (land, labour, and capital), or four (three plus industrial organisation). His treatment of organisation raised several issues of current importance. A short list includes claims that good organisation improves productive efficiency; the division of labour improves performance; the concentration of specialised industries in particular localities (*agglomeration* today) improves performance; large-scale production confers many advantages to business; and these advantages can be constrained or eliminated by government interference with the freedom of industry. All rely on good management, to which he devotes Chapter XII, and which has re-emerged in the twenty-first-century literature, as I explore in section "Drivers of Productivity Change".

The third giant, A. C. Pigou (1920 [1960]), made no observations of Antiquity, and it is hard to find reference to productivity in nearly 900 pages of his treatise on welfare. Nonetheless, he contributed much of substance directly related to this survey. Pigou characterised the national dividend as things purchased with money income, and considered it as a component of economic welfare, that part of total welfare that can be measured with money. He treated increases in the dividend as enhancing economic welfare, provided that the size of the dividend accruing to the poor is not thereby diminished. However, he was careful to state that a reduction in the inequality of the distribution of the dividend would increase economic welfare only under certain conditions concerning the definition of inequality. He then devoted the entire Part IV to the distribution of the dividend. In Chapter IV he argued that all inventions, both product and process, must increase the dividend, although because they "…may change the parts played by capital and labour in production in such a way as to make labour less valuable relatively to capital…", they do not necessarily increase the share of the dividend accruing to labour, which he loosely associated with the poor. This argument predates the current decline in labour's share of national income in many economies.

Pigou's distinction of the national dividend from economic welfare provides a good backdrop for the concurrent expression of social concerns surveyed in section "Social Concerns", and for the later revival of the issue in the inclusive green growth movement surveyed in section "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth". His analysis of the distributional impacts of invention underpins the never-ending "machinery question" covered in sections "Social Concerns" and "Productivity and the Pandemic Depression". His analysis of what we now call "Pigouvian" taxes and subsidies to increase the dividend by correcting for resource misallocation has found prominent application to environmental issues.

A Mere Century Ago: Accumulating Methods and Evidence Amidst Emerging Social Concerns

Four significant, and related, developments occurred during the last century. The first is the creation of index numbers, which was fundamental to gathering evidence. The second and third reflect an interest in gathering evidence on productivity trends, amidst growing concerns about the association of productivity growth with social progress. The fourth was the unfortunately brief appearance of the European Productivity Agency and its house journal *Productivity Measurement Review*, which publicised productivity relationships at the level of the individual firm, and even its plants and its production processes.

Methods

Evidence on aggregate productivity change in section "Evidence" is based on the use of index numbers to track productivity developments through time. The following brief overview of index numbers and their origins owes much to Diewert (1993) and Balk (2008), both of whom provide references to authors and their indices, after converting their price indices to quantity indices.

In 1823. Lowe proposed a fixed-base quantity index of the form $Q = \left(\frac{rq^1}{rq^0}\right)$, where q^1 and q^0 are quantities in successive time periods, but he left the time period for the price weight r unspecified. Subsequently, Laspeyres specified base period price weights $r = r^0$, so that $Q_L = \left(\frac{r^0q^1}{r^0q^0}\right)$, and Paasche specified comparison period price weights $r = r^1$, so that $Q_P = \left(\frac{r^1q^1}{r^1q^0}\right)$. Marshall and Edgeworth proposed arithmetic mean price weights $r = \bar{r} = ½(r^0 + r^1)$, with $Q_{\mathrm{ME}} = \left(\frac{\bar{r}q^1}{\bar{r}q^0}\right)$. Sidgwick and Bowley suggested the arithmetic mean of Q_L and Q_P, with $Q_{\mathrm{SB}} = ½\left[\left(\frac{r^0q^1}{r^0q^0}\right) + \left(\frac{r^1q^1}{r^1q^0}\right)\right]$, and Fisher recommended the geometric mean of Q_L and Q_P, with $Q_{\mathrm{F}} = \left[\left(\frac{r^0q^1}{r^0q^0}\right) \times \left(\frac{r^1q^1}{r^1q^0}\right)\right]^{1/2}$. Fisher called his index the "ideal" index, and his index does indeed perform best according to the test approach and the economic approach to evaluating the performance of index numbers; Balk (1995, 2008) and Diewert (in press a, b) provide authoritative treatments of the two approaches.

It is easy to build a productivity index from any of the quantity indices above. I follow tradition by defining a productivity index as an output quantity index divided by an input quantity index. For example, a Laspeyres productivity index is

$$\left(\frac{Y}{X}\right)_L = \frac{Y_L\left(y^1, y^0, p^0\right)}{X_L\left(x^1, x^0, w^0\right)} = \frac{\left(\frac{p^0y^1}{p^0y^0}\right)}{\left(\frac{w^0x^1}{w^0x^0}\right)},$$

a Paasche productivity index is written similarly, and a Fisher productivity index is

$$\left(\frac{Y}{X}\right)_F = \frac{Y_F\left(p^1, p^0, y^1, y^0\right)}{X_F\left(w^1, w^0, x^1, x^0\right)} = \left[\frac{Y_L\left(y^1, y^0, p^0\right)}{X_L\left(x^1, x^0, w^0\right)} \times \frac{Y_P\left(y^1, y^0, p^1\right)}{X_P\left(x^1, x^0, w^1\right)}\right]^{1/2}$$
$$= \left[\frac{\left(\frac{p^0y^1}{p^0y^0}\right) \times \left(\frac{p^1y^1}{p^1y^0}\right)}{\left(\frac{w^0x^1}{w^0x^0}\right) \times \left(\frac{w^1x^1}{w^1x^0}\right)}\right]^{1/2},$$

in which y^1 and y^0 are output quantities with prices p^1 and p^0, and x^1 and x^0 are input quantities with prices w^1 and w^0. The other quantity indices become productivity indices in the same fashion. An important feature of productivity indices is that they are empirical, functions of observable quantities and prices (in principle, although not always in practice), and can be calculated directly from the data without having to estimate anything. Most contributors I survey in section "Evidence" use index numbers to calculate productivity change.

In the above analysis, prices are used to weight quantity changes. However, Hamilton (1944) observed that historical price series dating back to the twelfth century "…are the oldest continuous objective economic data in existence", and contended that these data can reflect "…much better than can other attainable historical data changes in relative technological efficiency…" Subsequently a minority of writers, apparently unaware of Hamilton's contention, also have argued that price changes may provide useful measures of productivity change. Dayre (1951) and Fourastié (1951) both observed that real wages move proportionately with labour productivity, and H. S. Davis (1955; 29–30) claimed that "..productivity change and price change are in effect different sides of the same coin…" Fourastié (1951, 1957) stressed the "scientific importance and practical utility" of productivity measurement based on prices, which he called "indirect" productivity measurement, and he illustrated his point with detailed historical studies of trends in real wages (e.g., approximately 50 Indices of Change in the Level of Living During the First Century of Technical Progress in France [1830–1955]). Siegel (1952, 1955) was perhaps the first to specify an explicit price-based productivity index, essentially by reversing the roles of prices and quantities in the above analysis. He created a Laspeyres price-based productivity index

$$\left(\frac{W}{P}\right)_L = \frac{W_L(w^1, w^0, x^0)}{P_L(p^1, p^0, y^0)} = \frac{\left(\frac{x^0 w^1}{x^0 w^0}\right)}{\left(\frac{y^0 p^1}{y^0 p^0}\right)},$$

and a Paasche price-based productivity index is created similarly. The geometric mean of the two is a Fisher price-based productivity index

$$\left(\frac{W}{P}\right)_F = \frac{W_F(x^1, x^0, w^1, w^0)}{P_F(y^1, y^0, p^1, p^0)}$$

$$= \left[\frac{W_L(w^1, w^0, x^0)}{P_L(p^1, p^0, y^0)} \times \frac{W_P(w^1, w^0, x^1)}{P_P(p^1, p^0, y^1)} \right]^{1/2}$$

$$= \left[\frac{\left(\frac{x^0 w^1}{x^0 w^0}\right) \times \left(\frac{x^1 w^1}{x^1 w^0}\right)}{\left(\frac{y^0 p^1}{y^0 p^0}\right) \times \left(\frac{y^1 p^1}{y^1 p^0}\right)} \right]^{1/2}.$$

It is easy to show that $W_F/P_F = \Upsilon_F/X_F$ defined above if, and only if, $R^1/C^1 = R^0/C^0$, which requires Georgescu-Roegen's (1951) "return to the dollar" to remain constant through time. This equality holds in national accounts, in which the prices used to weight quantity changes are implicit deflators that convert nominal values to real values. This equality also provides the foundation for primal and dual growth accounting, pioneered by Jorgenson and Griliches (1967). However, errors of measurement in the national accounts or distorted prices in other contexts can cause the equality to fail, which implies that one index is more accurate than the other. This possibility motivated Hsieh (2002) to adopt a dual growth accounting approach to productivity measurement in East Asian countries, and he found overstated investment expenditure in Singapore the source of error in the primal growth accounting productivity estimate. Fernald and Neiman (2011) argued that measurement errors and distortions caused the two indices to diverge from one another, and from true productivity growth, in Singapore. Further theoretical and empirical research into price-based productivity measurement would add value.

Evidence

Early Estimates of Labour Productivity

Numerous early contributors to *Monthly Labor Review* reported on studies of productivity dispersion across establishments, and even across specific production processes within establishments that do not require index number techniques with which to aggregate process outputs. These studies, conducted under the auspices of the Bureau of Labor Statistics of the US Department of Labor, provided early illustrations of what can be achieved with cross-sectional samples collected at the establishment level. They are precursors to subsequent focused-sample productivity studies, dubbed "insider econometrics" by Ichniowski and Shaw (2012).

One of the earliest contributors was Squires (1917), who reported results of a study of labour productivity in the lumber industry. The study was based on data collected from 26 sawmills across more than 20 distinct manufacturing processes that stretched "…from tree to lumber pile" for a selected period of operation. In one particularly detailed investigation conducted over 11 representative logging establishments and five processes, it was possible to calculate labour productivity (output in board feet per man-hour) and unit labour cost (wages per board foot). Inter-establishment variation in labour productivity ranged from 5:1 to 50:1 across processes, and in unit labour cost ranged from 4:1 to 12:1 across processes. Squires attributed an unknown part of this dispersion to variation in the size of trees and in the dimension of lumber sawed, and to variation in methods of production and handling of the finished product. Squires' attribution illustrates an age-old challenge in productivity measurement—the role played by omitted variables in the search for the sources of measured productivity change.

Two decades later the story remained unchanged; only the setting was new. Stern (1939) reported results of a study of the boot and shoe industry from 1923 to 1936. The study reported labour productivity (number of pairs of shoes produced per man-hour) in 43 plants, 23 making men's shoes and 20 making women's shoes. In 1923, labour productivity varied by a factor of 4.3:1 in men's shoes and by a factor of 2.4:1 in women's shoes. By 1936, productivity dispersion increased to 5:1 for men's shoes and 4.3:1 for women's shoes. A distinctive feature of this study was its explanation for such wide and persistent inter-plant productivity dispersion. Among the likely sources cited were variation in management efficiency and in the skill and dexterity of individual operators, variation in the installation and use of specialised machinery economical only in large plants, variation in the rate of capacity utilisation of machines, and variation in shoe style ("particularly women's shoes"). Many of these sources reappear in modern studies of productivity dispersion and its persistence.

A notable feature of these two studies is that both were focused-sample inter-firm comparisons that did not require index number techniques. However, most studies of the period were aggregate time-series studies of productivity change, which did require index number techniques to aggregate variables. I survey a few of these studies, not primarily to recount their estimates of productivity change, but rather to illustrate different features of productivity measurement each study raises.

In his path-breaking study of US manufacturing industries over the period 1899–1914, Mills (1932) used a variant of Fisher's "ideal" index numbers to calculate changes in the physical volume of production $G_Y = 3.9\%$ pa, the number of wage earners $G_L = 2.2\%$ pa, and output per wage earner $G_{Y/L} = 1.7\%$ pa. Consequently, increases in labour productivity accounted for nearly half of output growth. Mills also calculated changes in the number of establishments $G_E = 1.1\%$ pa, and output per establishment $G_{Y/E} = 2.8\%$ pa, indicating that output growth exceeded growth in the number of establishments. He attributed these trends to technical and mechanical improvements, enhanced skills, increased technical efficiency (perhaps reflecting the influence of Taylor's scientific management), and to a trend toward large-scale production. He also calculated the same indices for 1913–1923 and 1923–1929.

Fabricant (1940) extended Mills' time series of US manufacturing industries to 1899–1937 and reoriented his focus from labour productivity to output per capita. He used a variant of the Marshall-Edgeworth index to calculate output growth $G_Y = 3.5\%$ pa (manufacturing output did not recover to its 1929 level until 1937), population growth $G_P = 1.4\%$ pa, and output per capita growth $G_{Y/P} = 2.1\%$ pa. He did not relate employment growth to population growth, and consistent with the title of his book he paid scant attention to productivity. Two years later, Fabricant (1942) extended his study period to 1899–1939 and reversed course by sharpening his focus to unit labour requirements, the reciprocal of labour productivity. He distinguished total employment from wage earners and incorporated declining trends in working hours per week for both. He calculated a decline of over 50% in the number of wage earners per unit of output, and a much stronger decline in wage earner hours per unit of output. He traced these declines to several causes, including automation, "novel and flexible" sources of power, giant factories, nationwide industrial networks, and revised methods of labour management, many of which remain relevant nearly a century later.

It is difficult to do justice to the breadth, depth, and historical coverage of C. Clark's *Conditions of Economic Progress* (Clark 1940, 2nd ed. 1951). In one exercise, he calculated labour productivity (real product per hour) for over 30 countries over varying long periods, 1800–1947 for the US, 1860–1947 for Great Britain, and 1789–1938 for France, with output measured by real national income expressed in International Units (i.e., USD, 1925–1934) and labour measured by hours worked. In a second exercise, he calculated labour productivity over varying periods for most of the same countries (including Palestine, Arabs, and Jews) in each of three sectors of the economy, primary [agriculture (e.g., the distributive cost of a kilo of peas in Brisbane, Australia, in 1939, retailing only), forestry and fisheries], manufacturing (excluding small scale), and tertiary (commerce, transport, services, and small-scale manufacturing). In both exercises, he found wide variation in productivity, both through time and across countries. Fourastié (1957; 97) praised Clark for the breadth of his research, writing that his 1940 book "…marked the beginning of a new era in economics, if only because of the systematic use made of statistical measurement, and the importance attached to the concept of long-term progress". He then criticised Clark for confusing value productivity with physical productivity, a recurring problem in productivity analysis explored in depth by Bartelsman and Wolf (2018).

Later Estimates of Total Productivity

Perhaps the first to generalise labour productivity to total productivity was George Stigler (1947), recipient of the 1982 Nobel Prize in Economic Sciences—members of this august group are honoured with given names—who calculated indices of labour productivity, capital productivity, and total productivity for a dozen US manufacturing industries over the period 1904–1937. He found wide variation in each productivity measure across industries and emphasised two features: the importance of total productivity, highlighted by the wide difference within each industry between trends in labour productivity and capital productivity; and the impact of the choice of base period, illustrated by modest (with one exception) differences in total productivity with 1937 weights and 1904 weights, a twist on the Paasche-Laspeyres spread.

Schmookler (1952) estimated changing efficiency of the aggregate US economy over the decadal period 1869–1878 to 1929–1938. He defined output as gross national product and input as the weighted sum of labour, land, capital, and enterprise, all expressed in 1929 prices. Enterprise was a novel input, constructed as a function of managerial labour and entrepreneurial capital. Total productivity grew at 1.36% pa (with labour measured in man-hours) or 0.92% pa (with labour measured in man-years), the difference highlighting a trend toward a declining work week. Productivity gains accounted for about half of growth in gross national product over the entire period.

Abramovitz (1956) also tracked the efficiency of the aggregate US economy through the updated decadal period 1869–1878 to 1944–1953. He defined output as real net national product per capita and real input as a weighted sum of labour (man-hours) and capital (land, structures, durable equipment, inventories, and net foreign claims) per capita. He estimated real output to have tripled, while real input increased by just 14%, and so growth in total productivity accounted for almost the entire increase in real output over seven decades. This result is somewhat surprising in light of previous findings based on similar data. It does however anticipate Solow's (1957) memorable 87½% productivity contribution. Abramovitz characterised total productivity equally memorably as "…the complex of little understood forces…" and as "…some sort of measure of our ignorance about the causes of economic growth…" (pp. 6, 11). This acknowledged ignorance has spawned a flood of research, some of which is surveyed in section "Drivers of Productivity Change".

Kendrick (1956) investigated productivity trends in the US economy over the period 1899–1953. He defined output as real private domestic product, labour as man-hours worked, and capital as the real value of land, plant, equipment, and inventories. Like several previous contributors, he acknowledged the inability to account for quality changes, and he compared total productivity growth with partial productivity growth associated with labour and capital, with the usual qualitative finding. Kendrick estimated total productivity growth of 1.7% pa, just over half the average rate of growth of real output of 3.3% pa. Soon thereafter, Kendrick (1961) added four years to his previous data and found essentially the same rate of total productivity growth of 1.7% pa, which continued to account for approximately half of output growth. However, for our purposes the two most interesting contributions of his 1961 study have been largely ignored, and warrant mention. The first was Kendrick's demonstration that the estimated rate of growth of the ratio of average total factor price to average product price is "identical" to the estimated rate of total productivity growth. This result, which holds only under the constant profitability condition, anticipates the development of price-based productivity indices, which I summarise in section "Methods". The second was Kendrick's use of information on input price trends to apportion the benefits of productivity growth to labour (99%) and capital (1%) during the period, a dramatic departure from current concerns about the recent decline in labour's share of national income. Kendrick's interest in the ability to use price changes to measure productivity change, and to examine the distribution of the fruits of productivity growth, set him apart from most previous and many subsequent writers.

Social Concerns

Concerns about the limitations of national income and related economic measures were expressed forcefully by Simon Kuznets (1934; 7), the 1971 recipient of the Nobel Prize in Economic Sciences. Kuznets argued that "…[t]he welfare of a country can… scarcely be inferred from a measurement of national income…", to which H. S. Davis (1947; vii) added "[n]o more important objective could be set…than that of increasing our knowledge of the conditions which stimulate and those which retard economic progress". These concerns feature prominently in Gordon's (2016) history of American growth, which lacks the expected qualifier "economic" in its title and is subtitled "The U.S. Standard of Living since the Civil War". His treatise claims that output measures miss the extent of revolutionary change from 1870 to 1940 and, to a lesser extent, since 1940, and documents massive improvements in the standard of living not incorporated in national income statistics.

The concept of social economic progress, social progress with an economic core that H. S. Davis (1955) contended was necessary, has its roots in the early twentieth century. Somewhat belatedly, it has inspired a twenty-first-century revival of the development of holistic approaches to productivity measurement, both micro (CSR, ESG, and related issues I do not cover in this survey) and macro (the OECD's inclusive green growth programme, which I survey in section "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth").

The Machinery Question

Early twentieth-century social concerns were perhaps first voiced in numerous contributions to the *Monthly Labor Review*. These contributions raised two issues of importance to labour, a core component of productivity analysis. The first chronicled injuries to, and threats to the health of, labour in industry. The second chronicled "technological unemployment", the displacement of labour by machinery, and the challenge of re-employment of displaced labour, a concern first raised well before the Great Depression, and most notably voiced during the Great Depression by Keynes (1931), who worried about the impact of technological unemployment on the economic possibilities for his grandchildren. The issue remains relevant to this day, in no small part because the productivity element of the issue has tended to remain in the background.

Some early writers acknowledged that technological improvements brought productivity gains that created displacement; to cite one example, a study of the displacement of labour by machinery in the glass industry (*Monthly Labor Review* 24:4, April 1927) found gains in labour productivity across several processes and products. In one extreme case of glass blowing, a process unchanged from that used in Egypt some 3,500 years ago, the introduction of new machinery raised labour productivity 41-fold, reduced employment, and altered the nature of employment from skilled glass blowers and unskilled child labour to mechanics and machine operators. Soon thereafter in the same *Review*, J. J. Davis (1927), US Secretary of Labor at the time, surveyed extant studies of the magnitude of labour productivity gains across numerous industries, and wondered "[w]hat are we doing with the men displaced?" In the case of telephone operators, it was young women who were displaced, in one of the largest automation shocks of the early twentieth century; this displacement and its labour market impacts, but not its productivity impacts, have been examined by Feigenbaum and Gross (2020). The widespread introduction of new technology brought both labour displacement and changing skill requirements, as well as productivity improvements, two features that remain relevant nearly a century later.

During and immediately following the Great Depression, many other writers understandably ignored the productivity gains created by mechanisation and concentrated on the possibilities for reabsorption of displaced labour by new and growing industries, in some instances reinforced by barriers to immigration. Lubin (1929a, 1929b), Myers (1929), and Clague and Couper (1931) conducted similar studies of technological unemployment, recording information such as the duration of unemployment, the source and destination industries of re-employed workers, the age distribution and geographic mobility of displaced workers, and their earnings distribution while displaced. Lubin noted that "newer" industries and trades were absorbing workers displaced from "older" industries and trades prior to the Great Depression, providing an early example of productivity-enhancing reallocation discussed in section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies". Jerome (1932, 1934) summarised these and many more studies of the time, and usefully distinguished productivity-enhancing from labour-displacing mechanisation, and provided a detailed discussion of the potential skill bias of each, providing an early example of complementarities discussed in section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies".

The machinery question was considered to have such broad public interest that *Scientific American* (1982) devoted an entire issue to "The Mechanization of Work". Berg (2010) has released a fascinating political economy history of the question.

More recently, *The Economist* (25 June 2016, 26 August 2017) has traced the job-*creating* dimension of technical progress back to the arrival of the first printed books in the 1470s, even before Ricardo raised the "machinery question" in the nineteenth century. Mokyr et al. (2015) have surveyed the long history of the machinery question and the resulting "technological anxiety" from the industrial revolution to the Great Depression. The machinery question has long been a contentious issue in agriculture, in which enormous productivity gains from mechanisation came simultaneously with comparably large labour displacement. The introduction of the mechanical cotton picker in the American south in the 1940s spawned labour displacement in the form of a wave of black migration to the northern industrial cities (Lemann [1991], who reported that "…picking a bale of cotton by machine cost the farmer $5.26 and picking it by hand cost him $39.41"). The question even led to an ultimately unsuccessful court case threatening agricultural mechanisation research, and hence agricultural productivity as well as agricultural employment, in the late twentieth century (Martin and Olmstead 1985, Los Angeles Times 1989).

The current literature on the machinery question is large and growing, with studies of automation, information technology, artificial intelligence, and robotics consistently finding both productivity gains and reduced employment in originating industries and exploring the implications for aggregate employment. Arntz et al. (2016) distinguished heterogeneous tasks within occupations from occupations themselves and concluded that automation is unlikely to destroy large numbers of jobs, with just 9% of OECD jobs at risk. Additionally, they argued that the introduction of new technologies may have two secondary effects, switching of tasks within occupations and additional job creation. Autor and Salomons (2018) expanded on the job creation possibilities, by distinguishing direct from indirect effects of automation. They found that automation enhances total productivity and reduces employment in originating industries, but that the direct employment losses are offset by indirect employment gains in upstream and downstream industries and by induced increases in aggregate demand through final demand and composition effects. In a series of closely related studies, Acemoglu and Restrepo

(2018, 2019, 2020) and Graetz and Michaels (2018) have examined the impacts of productivity-enhancing automation on employment. All combine a task-based framework with a decomposition strategy that allocates employment change to a direct displacement effect and a variety of indirect effects, including a reinstatement effect that captures the re-employment concerns of earlier writers. The indirect effects include the introduction of new technologies that create new tasks in which labour has a comparative advantage, increasing both productivity and employment. Empirical evidence is mixed, but a common finding is one of net productivity gains, small net employment gains or losses, and a strong skill bias to employment changes.

In a continuing series on the future of work, McKinsey & Company (multiple dates) optimistically predicted automation would boost global productivity growth by up to 1.2% pa, contribute to the solution of a range of societal challenges, and transform the nature of work. Autor et al. (2019) tempered McKinsey's optimism a bit, but only by conditioning similar predictions. The authors emphasised that not all productivity-enhancing technologies displace workers, and not all innovations that displace workers raise productivity, and they stressed the necessity of "…integrating technology with complementary innovations in work systems and management practices…" that magnify the productivity benefits of new and emerging technologies such as artificial intelligence and robotics. The emphasis on complementarities between new technology and work systems and management practices permeates current research, as exemplified by the OECD (2019), Gal et al. (2019) and Sorbe et al. (2019). These studies stressed the crucial role of management in harnessing new technology and suggest the hypothesis that new technologies can be managed more or less productively. Bloom et al. (2012) have tested this hypothesis, using two large micro panel data sets that enabled them to compare the performance of UK establishments owned by US multinationals, establishments owned by non-US multinationals and purely domestic establishments. They found establishments owned by US multinationals obtained higher productivity than either other group, a difference they attribute largely to superior US "people management" practices that enable US establishments to exploit IT more productively. I return to a more inclusive survey of the role of management in driving productivity gains in section "Drivers of Productivity Change", and to a discussion of artificial intelligence as a general-purpose technology and its likely impact on productivity growth in section "Productivity and the Pandemic Depression".

Social Economic Progress

More general social concerns centred on whether conventionally measured productivity gains were sufficient for achievement of a more inclusive objective of social economic progress (another old issue recently rediscovered, and briefly recounted in section "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth"). Fagan (1935) was an early proponent, defining economic progress as increased production accompanied by increased leisure, reduced natural resource depletion, and an "ever wider" distribution of productivity-generated wealth. C. Clark (1940; Chapter 1) proposed a similarly broad definition of economic progress, emphasising an increase of leisure, a "just" distribution of the fruits of productivity growth, the security of livelihoods, and a reduction in the "wasting" of natural assets such as minerals, timber, and the natural fertility of soils and pastures. He recognised that some of these objectives may be mutually inconsistent. H. S. Davis (1947) argued that productivity growth was necessary, but not sufficient, for economic progress. Among his additional requirements were an appropriate provision of seed capital for future production, relatively rapid re-employment of displaced labour and other resources, an increase in leisure, a "balanced" distribution of income, and an avoidance of wasteful use of natural resources. Spengler (1949) wrote of socio-economic growth, and distinguished economic *advance* from socio-economic *improvement*, citing exhaustion of non-reproducible natural resources as separating the two concepts and noting that soil-exhausting agricultural practices have been advanced as an explanation for the decline of ancient Rome. Writing in the same volume, J. M. Clark (1949) asked whether quantitative growth provided a fair index of real economic advance. In his reply to his own rhetorical question, Clark distinguished "sound and self-sustaining" growth from "unsound parasitic" growth, citing the wasting of the heritage of non-reproducible natural resources.

An obvious challenge arises, one that was not adequately addressed, and indeed largely ignored, by these early proponents of social economic progress. How does one define, and then measure, components such as a just distribution of income, adequate health and leisure, environmental degradation and exploitation of natural resources, or economic activity carried out in the household? Gary Becker, recipient of the 1992

Nobel Prize in Economic Sciences, was a founder of modern household economics, writing extensively on household consumption, production, and time allocation; he summarised his contributions to this and other fields in Becker (1993). Ahmad and Koh (2011) have begun the process of incorporating non-marketed household activity into the accounts, and Schreyer and Diewert (2014) have provided theoretical foundations for their incorporation. Income accountants around the world have established satellite accounts. However, integrating them with core accounts that would enable a systematic broadening of the conventional concept of total productivity remains work in progress. A consortium led by the United Nations has made valuable progress, with the 2008 System of National Accounts United Nations (2009) suggesting the structure of satellite accounts for tourism, the environment, health and unpaid household activity, which addresses some of the components of social economic progress. The United Nations also has produced the System of Environmental-Economic Accounting 2012 (United Nations 2014).

The Brief Flourishing of the European Productivity Agency

The European Productivity Agency was established by the Organisation for European Economic Cooperation in 1952, began operating in 1953, and ceased operations in 1961. During this time, it published *Productivity Measurement Review*, which contained summaries, often anonymous, of empirical productivity-related studies of two often overlapping sorts, with an objective of disseminating knowledge of best management practices. Both focused on productivity dispersion, reduction in which would enhance overall performance. The first consisted of inter-firm and inter-plant productivity comparisons, usually of labour productivity and always very detailed, reminiscent of those reported previously in *Monthly Labor Review*. One surveyed study conducted by the Netherlands Central Office of Statistics (*Productivity Measurement Review* 7, November 1956) illustrates the first sort. It tracked labour productivity dispersion among 12 Dutch bicycle manufacturers over six quarters during 1954–1955. For example, man-minutes of direct labour per unit of product were recorded across eight departments and ten operations. Inter-firm productivity dispersion was "striking", ranging from 1.7:1 (tyre-fitting) to 16.4:1 (front forks). A glaring omission from the comparison, acknowledged by the authors and a serious and continuing threat to many productivity comparisons, was a control for the type of bicycle manufactured. The second concerned variability of financial performance by examining cost structures and pyramids of financial ratios, the latter building on the duPont triangle approach to decomposing variation in return on assets. A surveyed study conducted by three national productivity institutes (*Productivity Measurement Review* Special Number October 1961) illustrates the second sort. It conducted an inter-firm comparison of the performance of 23 shoe-manufacturing firms in 1957. A total of 51 performance criteria were recorded, ranging from labour productivity and unit cost to the three components of the duPont triangle (return on assets = return on sales × asset turnover). Again, inter-firm performance dispersion was large. Labour productivity (pairs of shoes/hours worked) ranged from 0.20 to 1.79, and ROA ranged from −2.0 to implausible 31.3 and 38.4, illustrating a recurring problem: an unknown portion of observed dispersion in the 51 criteria was attributed to a failure, or inability, to account for variation in the type and quality of shoe produced.

Like the *Monthly Labor Review* before it, the *Productivity Measurement Review* contained cross-sectional focused-sample comparisons of micro units. Neither required index number techniques, but both could have benefited from the analytical techniques the next section surveys.

Converging to the Present: Analytical Foundations and Drivers

By the mid-twentieth century, we had a suite of index numbers with which to *calculate* productivity, and mathematical programming and econometric tools (not covered in this survey) had been developed with which to *estimate* productivity, but we lacked analytical models to structure and provide theoretical economic foundations for our estimation. Recent economic approaches to productivity measurement have gained in popularity and serve as a useful alternative to the older index number approach.

A virtue of both the index number approach in section "Methods" to productivity measurement and the analytical approaches to productivity measurement in section "Analytical Foundations of Productivity Measurement" is that both can be decomposed into a set of drivers of productivity change. A further advantage of the analytical approach is that each function can be decomposed, in ratio form, difference form or both, into a productivity change component and another component. This is a particularly useful property of parametric value functions, the decomposition of which can quantify the financial contribution of productivity change to change in alternative measures of business financial performance. In addition, the productivity change component can be further decomposed into various drivers of productivity change, yielding a two-stage decomposition of value change. This feature contrasts with productivity indices, which are not measures of financial performance and can only be decomposed into a set of drivers of productivity change. Balk et al. (2020) provide a detailed analysis of the methodology and the estimation procedures for both index number and analytical approaches to productivity measurement.

Analytical Foundations of Productivity Measurement

Analytical approaches to productivity measurement were developed well after the index number approach. Each analytical approach was developed initially within a cross-section context, which is incapable of estimating productivity change through time, although it can estimate productivity variation across firms. Extension to a time-series context with the objective of estimating productivity change came later. Many analytical models exist, and all can be adapted to a time-series context in which productivity change measurement is feasible. I discuss some of the more prominent models below.

Parametric Production Functions

Perhaps the first parametric function was a production function introduced by Cobb and Douglas (1928). Their function, still widely used empirically nearly a century after its publication, can be written as $y = AK^{\alpha}L^{1-\alpha}$, with y a single output, K and L capital and labour inputs, and $A > 0$ and $0 < \alpha < 1$ parameters to be estimated. A problem with this function, originally noted by Mendershausen (1938) is that it lacks a time dimension, making it unsuitable for the estimation of productivity change. Griliches (1996) cited Jan Tinbergen (1942), co-recipient of the first Nobel Prize in Economic Sciences in 1969, as the first to add such a shifter to the Cobb-Douglas production function, writing $P = a^{t}L^{\lambda}K^{1-\lambda}$. However, when he estimated this function using data on the US economy during 1919–1938, he judged the estimated coefficients unacceptable. Perhaps as a result, a blizzard of empirical applications soon followed. Among the first was Tintner (1946), who estimated the same production function with a time trend, which turned out to be plausible and highly significant, for the US economy during 1921–1941. Somewhat later Robert Solow (1957), the 1987 recipient of the Nobel Prize in Economic Sciences, then showed how to append a time shifter to *any* production function. Solow simply made A a function of time and wrote $y = A(t)f(K,L)$, from which it follows that

$$\ln y = \ln A(t) + \alpha \ln K + (1 - \alpha) \ln L$$

and

$$G_y = G_A + e_{yK} G_K + e_{yL} G_L,$$

with G indicating a growth rate and ε indicating a partial elasticity, both of which are unobserved. Under an assumption that inputs are efficiently allocated, partial elasticities are equal to cost shares, and this expression becomes

$$G_A = G_y - [S_K G_K + S_L G_L] = G_{y/L} - S_K G_{K/L},$$

with S indicating a cost share, which is observed. Solow calibrated this expression to aggregate US data over 1909–1949, from which he concluded that 87½% of growth in output per man-hour was attributable to G_A, which he called “technical change”. This left only 12½% to capital deepening. Another blizzard of empirical applications ensued, with most subsequent writers following a slightly different strategy, by treating the expression for G_y or $G_{y/L}$ as an equation to be parameterised and estimated. Either way, the exercise results in an inference about the famous “Solow residual” G_A, which provides an approximation to Υ/X, the productivity index obtained as the ratio of an output quantity index to an input quantity index. Notice three features of G_A: it does not require price information, whereas Υ/X does; it requires a single output, whereas Υ/X does not; and it supports a decomposition of observed output growth into the relative contributions of productivity change and input growth or, equivalently, a decomposition of labour productivity growth into the relative contributions of productivity growth and capital deepening.

Four observations are appropriate. First, the restriction to two inputs can be relaxed. Second, the production function above satisfies constant returns to scale, a restriction that can be relaxed by allowing the sum of the exponents to differ from unity. Third, the expression for ln$y$

is linear in ln$K$ and ln$L$ and provides a first-order approximation to the true but unknown technology. The approximation can be improved by adding squared and cross-product terms to the expression for ln$y$, providing a second-order approximation to the technology. To many scholars' embarrassment, it took 43 years for this seemingly obvious insight to be published, by Christensen et al. (1971)! Fourth, each of the first three observations apply to Cobb-Douglas type expressions for cost and revenue functions, respecting linear homogeneity of each in prices.

In principle, it is possible to extend the analysis from a single output to multiple outputs, by replacing a parametric production function with a parametric specification of a distance function introduced by Malmquist (1953) and Shephard (1953, 1970), simply by appending $A(t)$ to an input or output distance function. In practice, this is rarely if ever attempted, because a distance function has no natural variable to single out as the dependent variable in a regression exercise. Instead, two very different strategies are pursued. In one, a non-parametric distance function, which does not require a dependent variable, is specified and estimated using mathematical programming techniques. In the other, a natural dependent variable is created by aggregating either inputs or outputs and specifying a parametric value function such as a cost or revenue function. These two strategies are surveyed in sections "Non–parametric Distance Functions" and "Parametric and Non-parametric Value Functions".

Non-parametric Distance Functions

In this approach, a production function $y = f(x)$ with a single output and multiple inputs is replaced with a distance function $d(y,x)$ with multiple outputs $y \geq 0$ and multiple inputs $x \geq 0$. Define a production set $T = \{(y,x)$: y can produce $x\}$, an output set $P(x) = \{y$: y can be produced with $x\}$, and an input set $L(y) = \{x$: x can produce $y\}$. With an output-expanding orientation, the distance function becomes an output distance function $d_O(y, x) = \min\{\mu : y/\mu \in P(x)\}$, and with an input-conserving orientation the distance function becomes an input distance function $d_I(x, y) = \max\{\lambda : x/\lambda \in L(y)\}$. Rather than appending a time shifter $A(t)$ to each, a time indicator is attached to technology and variables, yielding $d_O^s(y^s, x^s)$ and $d_I^s(x^s, y^s)$, $s = 0{,}1$.

Distance functions have been applied to the measurement the efficiency of producers using an input vector to produce an output vector in a single activity in a single time period. This application was pioneered by Farrell

(1957) and has been generalised to data envelopment analysis (DEA) by Charnes et al. (1978). It can be further extended in two directions.

The first extension of distance functions is to the measurement of the efficiency of production in a multi-level activity in any number of time periods. The distinguishing feature of this extension is that resources must be allocated both within and among activities, which can occur in different time periods, a feature that characterises most modern production. This extension has several strands, the most popular of which is dynamic network DEA (NDEA). This rich literature, with its many empirical applications, is a direct descendent of the contributions to the analysis of the optimal allocation of resources of Leonid V. Kantorovich (1939) and Tjalling C. Koopmans (1951), co-recipients of the 1975 Nobel Prize in Economic Sciences, and Johansen (1972). As Peyrache and Silva (in press) chronicle, most contributors to the current NDEA and related literature ignore, or are unaware of, its rich heritage. The authors survey the analyses of the originators, the subsequent black box production models of Farrell and Shephard, and the current NDEA and related models.

The second extension of distance functions is to the measurement of productivity change. As in the case of index numbers, a productivity index is the ratio of an output quantity index to an input quantity index. Bjurek (1996) used distance functions to create Malmquist quantity indices, and from them a Malmquist productivity index; empirical applications appeared quickly. A Malmquist output quantity index comparing y^1 and y^0 is written as $Y(y^1, y^0, x) = d_O(y^1, x)/d_O(y^0, x)$, and a Malmquist input quantity index comparing x^1 and x^0 is written as $X(x^1, x^0, y) = d_I(x^1, y)/d_I(x^0, y)$. A period 0 output quantity index is $Y^0(y^1, y^0, x^0) = d_{\mathrm{O}}^0(y^1, x^0)/d_{\mathrm{O}}^0(y^0, x^0)$, a period 1 output quantity index is $Y^1(y^1, y^0, x^1) = d_o^1(y^1, x^1)/d_o^1(y^0, x^1)$, and a geometric mean output quantity index is $Y(y^1, y^0, x^1, x^0) = \left[Y^0(y^1, y^0, x^0) \times Y^1(y^1, y^0, x^1)\right]^{1/2}$. The two input quantity indices are written in the same way, and a geometric mean Malmquist productivity index is

$$\frac{Y(y^1, y^0, x^1, x^0)}{X(x^1, x^0, y^1, y^0)} = \left[\frac{Y^0(y^1, y^0, x^0)}{X^0(x^1, x^0, y^0)} \times \frac{Y^1(y^1, y^0, x^1)}{X^1(x^1, x^0, y^1)}\right]^{1/2}$$

and decomposes as

$$\frac{Y\left(y^1, y^0, x^1, x^0\right)}{X\left(x^1, x^0, y^1, y^0\right)} = \left[\frac{d_O^1\left(y^1, x^1\right)}{d_O^0\left(y^0, x^0\right)} / \frac{d_I^1\left(x^1, y^1\right)}{d_I^0\left(x^0, y^0\right)}\right]$$
$$\times \left(\left[\frac{d_O^0\left(y^1, x^1\right)}{d_O^1\left(y^1, x^1\right)} / \frac{d_I^0\left(x^1, y^1\right)}{d_I^1\left(x^1, y^1\right)}\right]\left[\frac{d_O^0\left(y^0, x^0\right)}{d_O^1\left(y^0, x^0\right)} / \frac{d_I^0\left(x^0, y^0\right)}{d_I^1\left(x^0, y^0\right)}\right]\right)^{1/2}$$
$$\times \left(\left[\frac{d_O^0\left(y^1, x^0\right)}{d_O^0\left(y^1, x^1\right)} / \frac{d_I^0\left(x^1, y^0\right)}{d_I^0\left(x^1, y^1\right)}\right]\left[\frac{d_O^1\left(y^0, x^0\right)}{d_O^1\left(y^0, x^1\right)} / \frac{d_I^1\left(x^0, y^0\right)}{d_I^1\left(x^0, y^1\right)}\right]\right)^{1/2}$$

The geometric mean productivity index has three drivers: the first component measures change in productive efficiency, as a production unit moves closer to or farther from best practice; the second component measures technical change, which expands or contracts production possibilities; and the third component measures exploitation of economies of size by moving along the production frontier that bounds production possibilities. A productivity index based on distance functions has three virtues: it allows multiple outputs and multiple inputs, it is independent of possibly mis-measured or missing prices, and it can support a narrative about the sources of productivity change. These three virtues have made this non-parametric productivity index an extremely popular vehicle for empirical productivity analysis. Russell (2018) provides a comprehensive overview of the Malmquist and other analytical productivity indices, and Aparicio et al. (2018) provide a recent empirical application to the provision of public education in Spain following the financial crisis, in which schools increased their productivity by raising academic achievement despite shrinking budgets.

Parametric and Non-parametric Value Functions

Edward S. Mason, a former President of the American Economic Association, wisely observed in his Preface to Dean (1941) that "…significant economic relationships may be derived from the accounting and operating data of a business firm". He then added, in a sign of the times, that the "…techniques here used, and at present available, are not so well suited to deal with the more complicated problems of a multi-product firm with changing methods of production". As we now know, the resort to value functions circumvents this problem by aggregating multiple variables to create a single variable, cost or unit cost or revenue or unit revenue or profit or some other indicator of business financial performance. This is the principal virtue of the use of parametric value functions to estimate productivity change, the ability to tailor them to the presumed objective of, and the constraints facing, a production unit.

I begin with a parametric cost function, which Dean estimated in short-run form, assuming a single output, fixed technology, fixed capital equipment, and fixed input prices, using monthly data for a leather belt shop. Soon thereafter, Nordin (1947) estimated a long-run cost function quadratic in a single output in an electric light and power plant during more than 500 shifts, and Lomax (1952) estimated a unit cost function log-linear in generating capacity and load factor for a sample of British steam electricity-generating plants. Neither study controlled for variation in input prices, although Lomax acknowledged the omission.

The use of a cost function is motivated by a business objective of keeping costs down, as at Ikea or Walmart, for example, or by an analyst's belief that productivity change has a resource-saving orientation. A cost function can be written as $wx = A(t)c(y,w)$, with y an output vector, w and x input price and quantity vectors, with $c(y,w) = \min_x\{wx: x \in L(y)\}$ a minimum cost function (or cost frontier) to be estimated, and t a time counter. A cost frontier has properties that must be imposed or tested, including monotonicity, concavity, and homogeneity of degree +1 in w. Following the same procedure as with a parametric production function yields

$$G_{wx} = \sum \varepsilon_y G_y + \sum \varepsilon_w G_w + G_A$$

with partial elasticicities ε and growth rates G. Cost change is driven by change in outputs produced, by change in input prices paid, with $\sum \varepsilon_w = 1$ to incorporate homogeneity of degree +1, and by cost-reducing technical change. Morrison (1992) illustrated decompositions of cost change along the lines of the above expression under a variety of scenarios relating to size economies and fixity of some inputs, with an empirical application to US, Canadian, and Japanese manufacturing.

This expression, especially with a single output, is popular, but it suffers from an untenable assumption that actual cost equals minimum cost. This assumption can be relaxed in either of two ways. One is by appending a non-negative component to a normally distributed error term to allow for cost inefficiency in addition to the usual sources of random noise, and to allow change in cost efficiency to drive observed cost change. This procedure generates a stochastic cost frontier model, and a stochastic production frontier model is created in the same way, with a two-part error term; Sickles and Zelenyuk (2019) provide details and a guide to a burgeoning literature.

In an alternative approach, let $C^0 = w^0x^0$ and $C^1 = w^1x^1$ denote observed cost in periods 0 and 1, with w and x denoting vectors of input prices and quantities, and let $wx \geqq c(y, w)$, with $c(y,w)$ a minimum cost function (or cost frontier) to be estimated. Change in observed cost from period 0 to period 1 is expressed in ratio form as

$$\frac{C^1}{C^0} = \frac{w^1x^1}{w^0x^0} = \frac{c^0(y^0, w^1)}{c^0(y^0, w^0)} \times \left[\frac{w^1x^1/c^1(y^1, w^1)}{w^0x^0/c^0(y^0, w^0)} \times \frac{c^1(y^1, w^1)}{c^0(y^1, w^1)} \times \frac{c^0(y^1, w^1)}{c^0(y^0, w^1)}\right]$$

which identifies two drivers of cost change, input price change (the first term on the right side, in which only the input price vector changes) and productivity change, which itself is the product of three components, change in cost efficiency, change in technology, and change in size (a combination of economies of scale and economies of diversification). Combining an estimate of $c(y,w)$ with observed cost enables one to distinguish input price change from productivity change as drivers of cost change, and also generates a story about the impact on cost change of the three components of productivity change. The introduction of efficiency change as an independent driver of productivity change in distance functions and value functions has enabled analysts to investigate an important managerial and public policy challenge, the minimisation of productivity-sapping and costly inefficiency in production.

It is straightforward to convert this expression from ratio form to difference form, which may appeal to managers comfortable thinking in terms of monetary values. To illustrate, cost change is expressed in difference form as

$$\begin{aligned} C^1 - C^0 &= w^1x^1 - w^0x^0 \\ &= \left[c^0(y^0, w^1) - c^0(y^0, w^0)\right] \\ &\quad + \left[w^1x^1 - c^1(y^1, w^1)\right] - \left[w^0x^0 - c^0(y^0, w^0)\right] \\ &\quad + \left[c^1(y^1, w^1) - c^0(y^1, w^1)\right] \end{aligned}$$

$$+\left[c^0\left(y^1, w^1\right) - c^0\left(y^0, w^1\right)\right]$$

which decomposes observed cost change into the sum of an input price effect and a productivity effect, which in turn decomposes into a cost efficiency change effect, a technology effect, and a size effect, each expressed in monetary terms. Each effect is measured in monetary terms. Both ratio and difference forms of cost change are expressed from a period 0 perspective. It is easy to express both from a period 1 perspective, and then to take the geometric mean of the two expressions. Grifell-Tatjé and Lovell (2003) applied the difference form decomposition, with managers representing observed cost and consultants representing minimum cost, in a sample of Spanish electricity distributors.

In principle, it is possible to use similar approaches to identify the contribution of productivity change to change in unit cost, which Bliss (1923) advocated for two purposes: to evaluate business financial performance, and to inform product pricing decisions. Gold (1971) decomposed unit cost change into changes associated with each input used in the production process and applied the decomposition to US iron and steel manufacturing during 1899–1939. He recognised the challenge of defining a "unit" of output in a multi-output setting, which he addressed by expensing profit and creating a "non-existent composite product", defined as total revenue. He found an output price index to have increased by less than increases in two of three input quantities, which he attributed to materials- and labour-saving improvements in technology. An alternative approach, which he also explored, is to define and decompose unit cost for each output separately and conduct a productivity analysis for each output, but this approach requires cost allocation. Kendrick and Grossman (1980) calculated unit cost in the US nonfinancial corporate business sector during 1948–1976, using an index of gross product originating as the unit of output. They found unit cost growth of 2.92% pa, with an average factor price increase of 5.16% pa partially offset by an increase in total productivity of 2.18% pa. For both Gold, and Kendrick and Grossman, productivity growth kept unit cost down by offsetting the upward pressure brought by input price increases. Grifell-Tatjé and Lovell (2015; Chapter 7) provide detailed treatments of decompositions of cost change, unit cost change, and unit labour cost change, with references to empirical applications.

Some companies, Netflix and Nvidia for example, seek rapid growth. The same three cost-oriented approaches apply under a business objective of pursuing maximum revenue growth, or when the analyst believes productivity change is output-augmenting. Observed data become (x,p,y), with p an output price vector, and the maximum revenue function $r(x, p) = \max_y\{py{:}y \in P(x)\} \geqq py$ must be estimated. The revenue function also satisfies properties, including monotonicity, convexity, and homogeneity of degree +1 in p, which must be imposed or tested. Approaches to the measurement of output-expanding productivity change are directly analogous to those for the measurement of input-saving productivity change above, with py replacing wx and $r(x,p)$ replacing $c(y,w)$. However, empirical applications are scarce.

Pursuit of maximum revenue may be constrained, in at least three interesting ways, each of which is easily motivated by observed business practices. In the private sector, branch or division or regional managers receive operating budgets and are assigned the objective of maximising revenue. In the 1920s management at duPont and General Motors had to decide how to allocate scarce investment and other resources across product lines and among plants. Their criterion was maximising return on those assets; Chandler (1962) and Johnson (1975, 1978) recount the history. In the public sector, it is the pursuit of maximum output that may be constrained, as when agency managers receive operating budgets and are assigned an objective of maximising output, usually in the form of service provision. In both cases, the input vector is no longer fixed, replaced by a fixed budget constraint of the form $wx \leqq B$, and managers are free to choose an input vector that maximises output subject to $x(w/B) \leqq 1$. The idea originated with Shephard (1974), and has found frequent application in the public sector, to the measurement of the performance of hospitals, for which performance is some measure of health outcomes such as QALYs or DALYs, and to the performance of schools or universities, for which performance is some measure of student outcomes such as standardised test scores or employment and income statistics. Staiger (2020) has noted that structural similarities support the use of similar productivity measurement techniques in the two sectors. Blank and Merkies (2004) have applied the budget constrained output maximisation model to Dutch hospitals, and Grosskopf et al. (1999) have applied the model to US school districts.

In a third scenario imperfect market competition allows empirical application of the old Berle and Means (1932) conjecture of the separation of ownership from management control. This separation gives managers discretion to pursue their own objectives, which are best served by maximising sales revenue. Owners cannot be ignored, however, and profit becomes a constraint rather than an objective. Here an analytical model that is structurally similar to the private sector model of Shephard is appropriate, with the maximum budget constraint replaced with a minimum profit constraint. Baumol (1959) proposed this model of profit-constrained sales revenue maximisation in a cross-section context, but it is easily adapted to a time-series context.

The standard textbook objective, if not the current socially responsible objective, of a business firm is profit maximisation, and it is enlightening to derive a profit maximisation model along lines similar to those for cost minimisation. Let profit in periods 0 and 1 be given by $\pi^s = p^s y^s - w^s x^s > reqless\, 0$, $s = 0{,}1$. Then profit change becomes

$$\pi^1 - \pi^0 = \left[p^0\left(y^1 - y^0\right) - w^0\left(x^1 - x^0\right)\right] + \left[y^1\left(p^1 - p^0\right) - x^1\left(w^1 - w^0\right)\right],$$

in which the first term on the right side is a Laspeyres quantity effect and the second term is a Paasche price effect, both in difference form. A few manipulations of both effects leads to the following decomposition of profit change

$$\pi^1 - \pi^0 = \left\{w^0 x^1\left[\left(\frac{Y_L}{X_L} - 1\right)\right] + \pi^0[Y_L - 1]\right\} + \left\{y^1 p^0\left[\left(\frac{P_P}{W_P} - 1\right)\right] + \pi^1\left[1 - W_P^{-1}\right]\right\}.$$

The first term on the right side decomposes the Laspeyres quantity effect into a productivity effect that converts a Laspeyres productivity index Υ_L/X_L into a monetary value, the productivity bonus, and a Laspeyres margin effect that attaches value to output expansion or contraction $Y_L > reqless\, 1$ with nonzero base period profit. The second term decomposes the Paasche price effect into a price recovery effect that converts a Paasche price recovery index P_P/W_P into a monetary value, and a Paasche margin effect that values input price increase or decline $W_p > reqless\,1$ with nonzero comparison period profit. This expression, which is easy to implement empirically, highlights the fact that productivity growth is not the sole source of improved financial performance; price recovery may be equally or more important, depending on market conditions. Grifell-Tatjé et al. (2018b) use this model to study the business foundations of social economic progress.

The New South Wales Treasury (1999), inspired by Eldor and Sudit (1981) and Miller (1984) in the business literature, implemented the profit change model above, which it called Profit Composition Analysis, to separate the productivity performance from the price recovery performance of government-owned businesses, particularly those possessing market power and subject to price regulation. Villegas et al. (2020) have applied the profit change decomposition model to the English and Welsh water and sewerage industry.

Drivers of Productivity Change

Now that we know how to model and estimate productivity change, we are equipped to delve into the factors that drive or impede it. Many drivers have been identified, perhaps the most significant of which appear below.

Quality Change

Several previously cited writers have lamented their inability to adjust variables for changes in their quality. More than other drivers, quality change is primarily a measurement issue; resolving the issue enables one to disentangle the separate contributions of quantity change and quality change to total change, of inputs, outputs, and external ("non-discretionary") variables characterising the environment in which production takes place. Denison (1962) provided an illustration of what can be achieved with good data. He calculated growth rates of real national income, labour, land, and capital over varying periods in the US. For 1929–1957 he calculated $G_Y = 2.93\%$ pa, $G_L = 1.57\%$ pa, $G_K = 0.43\%$ pa, zero growth for land, and total productivity growth $G_{Y/X} = 0.93\%$ pa. He attributed growth in the labour input to growth in employment, with $G_E = 1.00\%$ pa, and growth in several quality indicators that raised labour's contribution, including hours, education, experience, and changes in the age-sex composition of the labour force. He decomposed capital's contribution, but not to quality and quantity change, and he made no adjustment to real national income. He did, however, decompose productivity growth into seven components, including restrictions against optimum use of resources, sectoral shift from agriculture, and the primary component, advance of knowledge.

Denison (1974) subsequently revised these figures, the main revisions being increases in G_Y and $G_{Y/X}$ and a corresponding reduction in G_L. He also added a new series covering the faster growth period 1950–1962, and a longer 1929–1969 period for the non-residential business sector. In the interim, an extended debate ensued in the Survey of Current Business (1972), with Denison pitting his relatively large contribution of productivity growth to US output growth against a much smaller contribution estimated by Jorgenson and Griliches (1967). The debate, though instructive, generated limited convergence of views of the relative contribution of productivity growth to US economic growth.

Econometricians treat quality change as a type of specification error; Griliches (1957) illustrated the econometric issues involved. The OECD Productivity Manual, OECD (2001) and International Labour Organization (2020) have treated the theory and empirical adjustment of quantities and prices for quality change in great detail and have provided guidelines on the measurement of and adjustments to output, labour input, and capital input, and the measurement and interpretation of productivity in the presence of compositional effects. In contrast to adjusting outputs and inputs for compositional change, the challenge of incorporating features of the external operating environment into a productivity analysis has received considerable attention in the literature, but relatively little attention has been paid to how to measure it.

Technology

Technological drivers of productivity growth, including technical progress, efficiency change, and the exploitation of economies of size, made their first appearance in section "Analytical Foundations of Productivity Measurement". The empirical application of distance functions and cost functions to implement these decompositions is spread widely across private and public sectors. However, one driver is concealed in this tripartite decomposition. Economies of size is a generic term, encompassing the familiar radial notion of economies of scale with a less common notion of economies of diversification, a non-radial concept. It is important to distinguish the two, because firms tend to grow by altering the proportions of outputs they produce, by diversifying their product range, or even by specialising production in a single product, and the cost of expansion can be sensitive to the direction of output growth.

The economic analysis of economies of diversification is not new. Marshall (1890) devoted Book V, Chapters VI and VII to cost allocation in a multi-product firm, under the heading of "joint supply", against an institutional backdrop of the repeal of the Corn Laws. Penrose (1959) devoted Chapters VI and VII to economies of size and diversification, noting that in an environment of changing technology and tastes, or in the presence of temporary fluctuations in demand (e.g., seasonal), a firm can make more profitable use of its resources by spreading production over a variety of products. She also anticipated the subsequent analytical literature by exploring the trade-off between the sacrifice of scale economies in specialised production and the gain in cost complementarities from diversified production.

The analytical foundations of the economies of diversification are relatively recent, and are based on a multi-product cost function, or frontier, and most details are available for a limited sort of diversification economies named economies of scope in Panzar and Willig (1981) and for diversification economies more generally in Baumol et al. (1982). Applications are numerous, especially in the provision of multiple financial services; see for example Pulley and Braunstein (1992) and Cummins et al. (2010) among many others. Growitsch and Wetzel (2009) test for economies of scope in European railways, and de Roest et al. (2018) test for economies of diversification in EU agriculture. Such studies generally quantify the cost-oriented benefits of diversification, but there is a revenue side to business success as well, and the business world is littered with costly diversification failures—think of the Ford Edsel, the Sony Betamax, and the Apple Newton for example.

Organisation

I treat organisation as the role of management in enhancing business performance. I define management broadly to include both those who direct individual businesses and those who direct an aggregate economy through monetary, fiscal, trade, and regulatory policies; in the latter case management corresponds to the "helmsmen" of Koopmans (1951). The study of organisation has a rich history and a lively current literature, most of which is directed at the business enterprise rather than the aggregate economy.

Walker (1887) clearly understood the importance of management, calling it the source of business profits. This profit, which he called surplus, "…represents that which he is able to produce over and above what an employer of the lowest industrial grade can produce with equal amounts of labour and capital. In other words, this surplus is of his own creation, produced wholly by that business ability which raises him above and distinguishes him from the employers of what may be called the no-profits class". Among the components of business ability, Walker mentioned administrative and executive ability, including the ability to avoid waste, and the ability to meet changing market demands quickly. Alfred Marshall (1887), whose 1879 *Economics of Industry* Walker had favourably cited, wrote of the allocation of the surplus generated by superior management to rent and profit.

Later Marshall (1890; Book IV, Chapter XII) made another contribution to the literature on business management, combining and expanding on his and Walker's earlier writings by defining the *functions* of management. He regarded "business men" as a highly skilled industrial grade who undertake risks, bring together capital and labour, engineer the business, and superintend its minor details. The supply price of business men had three components: the supply price of capital, the supply price of business ability and energy, and the supply price of organisation that brings the first two together.

The distinguished management consultant Drucker (1954; 71) had insights that might have guided much subsequent work on management and productivity. He focused on the *quality* of management, claiming that "…the only thing that differentiates one business from another … is the quality of its management …" He continued by contending that the only way to measure managerial quality is by means of a "…measurement of productivity that shows how well resources are utilized and how much they yield". He defined the yield of utilised resources in terms of meeting multiple, often conflicting objectives: "There are few things that distinguish competent from incompetent management quite as sharply as the performance in balancing objectives". One of his objectives is public responsibility; see sections "Social Concerns" and "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth".

Penrose (1959) devoted Chapters III–V to entrepreneurs and managers, and treated managerial services as an essential input to a production process [i.e., $f(K,L,0) = 0$]. Management creates an inescapable limit to the growth a firm can achieve in any given period.

This managerial limit can, however, be relaxed through changes in external conditions such as the state of knowledge or the state of the arts. This association of a managerial limit with the frontier of production possibilities, and the potential for increasing production possibilities through external improvements, foreshadows a large literature on productivity gaps, distance to frontier, and catching up and falling behind considered in section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies".

But first we need to address the challenge of incorporating management into an analytical approach to productivity measurement treated in section "Analytical Foundations of Productivity Measurement". Early efforts of Hoch (1955, 1962), who called it "entrepreneurial capacity", and Mundlak (1961) and Massell (1967), who called its omission "management bias", showed one way of incorporating management into a parametric representation of production technology. Rather than including it as an input, since they did not pretend to know what it was or how to measure it, they estimated it as a firm effect $y = A_i f(K,L)$, with i indexing firms. All three were able to reject the null hypothesis of no firm effect, which signals the presence of productivity dispersion surveyed in section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies". Mefford (1986) was able to take a different approach, because he had a proxy for plant management in a large multi-plant international business. He constructed an index M of the rankings of managers' plants on three performance criteria: output goal attainment, budget over- or under-fulfilment, and output quality, asserting that these criteria are "…the major evaluative factors on which most industrial managers are judged and corporate management confirms this to be the case in this firm". Having this information, unavailable to Hoch, Mundlak, and Massell, enabled him to incorporate M as an input in a production function regression $y = f(K,L,M)$. Mefford found a statistically significant positive elasticity on M, implying a positive relationship between his proxy for management performance and output. The econometric issue involved is exactly what Mundlak and Massell called it, management bias. Griliches (1957) illustrated the econometric issue and explained the circumstances under which the omission of management does or does not bias estimates of the contributions of the included variables.

Nearly a half century after Hoch, Mundlak, and Massell wrote of management as a firm effect, Bloom, Van Reenen, and colleagues embarked on a research project that remains ongoing. They created a management score as a function of 18 key indicators of the quality of management practices grouped into four areas: shop floor operations, performance monitoring, target setting, and provision of incentives. They also proposed a pair of indicators that potentially drive the quality of management practices, the extent of product market competition and the nature of business ownership. They then developed five indicators of business economic and financial performance that potentially result from the quality of management practices, including total productivity, profitability (ROCE), Tobin's Q, sales growth, and survival. They also amassed large and growing inter-sectoral international databases with which to test hypotheses on the drivers and consequences of variation in the quality of management practices. Their methodology, data, and findings appear in a continuing series of publications beginning with Bloom and Van Reenen (2007).

Scur et al. (2021) recount the history of The World Management Survey, the database used by many contributors to the literature exploring the impact of management on business performance. In their initial study of 732 manufacturing firms in Europe and the US, they found the distribution of management practices to be dispersed and asymmetric, with relatively large variance and negative skewness. From this, they reached a number of conclusions: three different measures of the degree of product market competition are significantly and positively associated with better management practices; family ownership by itself is unrelated to the quality of management practices, but family ownership combined with primogeniture lowers the quality of management practices; a long tail of badly managed firms occupied primarily by primogeniture family firms operating in markets with low competition; significant variation in the quality of management practices across countries and a much larger variation within countries; and perhaps most significantly, the quality of management practices is strongly associated with all five indicators of firm performance.

Subsequent studies have been based on larger samples of up to 11,000 firms, often within sectors such as healthcare and education or within countries such as India, and frequently with different variables such as the human capital of managers. Thus Bloom et al. (2009) explored the relationship among work-life balance variables, management practices, and productivity. Using a sample of over 700 firms in Europe and the US, they found that, once management practices are included, work-life balance variables have no independent impact on productivity, contradicting the claims of some that higher productivity comes at the cost of work-life balance. Bloom et al. (2010) expanded the list of business performance indicators. They used a sample of over 300 manufacturing firms in the UK to explore the relationship between the quality of management practices and economic and environmental performance. They found better-managed firms to be significantly more productive, but not at the expense of the environment. Better-managed firms were also significantly less energy-intensive, and therefore generated fewer greenhouse gas emissions. Bender et al. (2018) used a sample of 365 medium-sized German manufacturing plants to investigate the role of a subset of management practices directed toward human resource management (HRM). They found that plants with higher management practice scores had above average worker skills gained by augmenting human capital through selective hiring and attrition. They also found that the human capital of management made a larger contribution to productivity than the human capital of the labour it employs. Later Bloom et al. (2019) expanded the list of drivers of variation in the quality of management practices, and they introduced additional drivers of economic performance. They used samples of varying sizes of US manufacturing plants and found "enormous" dispersion of management practices. Variation in the quality of management practices accounted for a greater share of variation in productivity, growth, and survival than did variation in R&D expenditure, ICT per employee, or employee human capital. They found two new drivers of the quality of management practices, the business environment as proxied by right-to-work laws that increase labour market competition and the arrival of "Million Dollar Plants" that generate learning spillovers. Lemos et al. (2021) decomposed management practices into operations management practices and people management practices in a sample of public and private schools in Andhra Pradesh, India. They found private schools to be better managed on both practices, although people management skills were the primary drivers of teaching quality and student value added. They also found that more effective private school teachers received significantly higher wages than less effective public school teachers. Bloom and colleagues have published many other studies concerning the quality of management practices, but this sample provides the essence of their contribution: when studying productivity, management matters.

The work of Bloom and colleagues has motivated a vast amount of related research. Adhvaryu et al. (in press a) unbundled the management practices index to focus on its people management skills elements and found that these skills increase productivity but not management pay. Adhvaryu et al. (in press b) studied management performance and productivity following air pollution shocks and found that managers with people management skills were better able to mitigate the adverse impact of air pollution shocks on productivity. Hoffman and Tadelis (2021) focused on managers' people management skills in a large high-tech firm; they found these skills to reduce attrition of those employees management wants to retain, and to benefit managers having these skills with higher promotion rates and, contrary to Adhvaryu et al. (in press a), larger salary increases. While much work inspired by that of Bloom and colleagues has examined the impact of the quality of management *practices* on organisational performance, Alexopoulos and Tombe (2012) expanded on the management *technologies* perspective on the relationship between HR and productivity introduced by Bloom and Van Reenen (2011) to examine the impact of innovations in managerial technologies on performance. Examples of novel managerial technologies include Taylor's scientific management, just in time, and total quality management. The authors used aggregate US data to examine the impact of 16 such indicators on productivity, and among their results, they found that managerial technology shocks raised productivity more rapidly than non-managerial technology changes do.

Reversing the usual strategy, Cai and Wang (2020) studied worker evaluation of management and found that evaluation of management by teams of workers improved the relationship between management and employees, reduced employee turnover, and increased team productivity. Gosnell et al. (2020) also studied the evaluation of management performance, in this case airline captains, and found performance monitoring and feedback, target setting, and prosocial incentive provision improved airline productivity, defined as a function of fuel use, time delay, and safety. Sickles et al. (2021) applied production theory to an unbalanced panel of 505 medium-size manufacturing firms in Europe and the US. They found the shadow price of management to exceed average management compensation, and therefore management to be relatively under-utilised, with both inequalities shrinking over time.

I conclude this discussion with the perhaps obvious observation that the people management skills elements of the 18 management practices identified by Bloom and colleagues are becoming increasingly valuable, but this observation is not new. Long ago, Drucker (1954) bemoaned the lack of attention paid to the HR component of management. He observed that "…we know too little about it so far, operate largely by superstitions, omens and slogans rather than by knowledge. To think through the problems in this area and to arrive at meaningful measurements is one of the great challenges to management". The challenge and the significance of the HR function remain, and over a half century later *The Economist* (26 March 2020) has observed that HR heads' desks are moving closer to the corner office. The academic literature and the business press both are stressing the significance of the HR element of management practices, and this emphasis is spurring interest in the role of HR practices and IT adoption, and complementarities between the two, as drivers of business performance, both productive and financial. Important contributions have been made by Black and Lynch (2001), Ichniowski and Shaw (2012), Bartel et al. (2007) and Lazear et al. (2015) among many others, most of whom emphasise the significance of complementarities between HR practices and IT adoption. Benner (2018) explored the importance of complementarities between management practices and IT adoption and use for business productivity, citing one hypothetical example of management's choice between allocating scarce resources to producing Blackberrys more productively or to creating the iPhone.

Institutions

No business operates in a vacuum. Rather, business operating environments are characterised by an institutional framework. Douglass North (1990), co-recipient of the 1993 Nobel Prize in Economic Sciences, described the framework as "…the rules of the game in a society or, more formally, …the humanly devised constraints that shape human interaction". As such, institutions influence the economic and financial performance of business, and therefore aggregate productivity performance. North illustrated the important role played by institutions in influencing aggregate economic performance by tracing the contrasting institutional histories of England and Spain from the beginning of the sixteenth century through their downstream consequences for the divergent twentieth-century economic performances of the US and Latin America.

In an influential study, Hall and Jones (1999) regressed output per worker against physical and human capital per worker for a 1988 cross-section of 127 countries, and found a large unexplained residual, suggesting that something else must influence inter-country variation in labour productivity. This large measure of their ignorance led them to develop an index of nations' social infrastructure, the institutions, and government policies that define their business operating environments. Further empirical analysis led them to the conclusion that variation in countries' productivity performance was determined primarily by variation in their social infrastructure. Soon thereafter Easterly and Levine (2001) studied time-series data for OECD, Latin American, and East Asian countries, and found with few exceptions factor accumulation to account for less than 2/3 of GDP growth, once again leaving a large unexplained residual. Citing numerous previous studies, they tested the hypothesis that "national policies" influence productivity growth, using policies including openness to trade, inflation, government size, financial development, and a black-market premium. They found all but inflation to exert significant influences on productivity.

Numerous recent studies have used the latest available version of the World Bank Group (2020b) *Doing Business* data to examine the impacts of institutions on business performance; the Group claims nearly 4,000 peer-reviewed articles and over 10,000 working papers have used the data since 2003. The data include indicators of difficulties involved in opening a business, getting a location, accessing finance, dealing with day-to-day operations, and operating in a secure business environment, from which the Group has constructed an aggregate index of the ease of doing business, currently for 190 countries (New Zealand ranks first). Each of the following studies used these data to emphasise particular institutions, either singly or jointly. Barseghyan (2008) and Barseghyan and DiCecio (2011) added to the list of institutional variables an index of entry barriers, which include costs an entrepreneur incurs for starting a new firm, building a physical location and meeting minimum capital requirements. They found that these entry barriers reduced labour productivity by reducing the total productivity residual and created a misallocation of resources that allowed unproductive firms to operate. Moscoso Boedo and Mukoyama (2012) investigated the effects of entry and exit (firing) costs on income and productivity across countries. They found that both costs reduced aggregate productivity, but through different channels. Entry costs reduced aggregate productivity by reducing both entry and exit, thereby keeping low productivity establishments in operation, while firing costs reduced aggregate productivity by hindering the reallocation of labour from low productivity establishments to high productivity establishments. D'Erasmo and Moscoso Boedo (2012) examined the impact on aggregate productivity of costly entry from the less productive informal sector, which looms large in developing countries, to the more productive formal sector with its better institutions such as debt enforcement mechanisms. They found that countries with low degrees of debt enforcement and high entry costs into the formal sector were characterised by low allocative efficiency and low aggregate productivity. Bergoeing et al. (2016) explored complementarities among reforms to entry and exit barriers. They found that only comprehensive reforms combining loosening of both entry and exit barriers had substantial effects on output growth and aggregate productivity.

Égert (2016) summarised OECD findings to date on the impact of product and labour market regulations and the quality of other institutions on aggregate total productivity for a panel of 34 OECD countries over a 30-year period. Among the findings were (i) active labour market policies, employment protection legislation, R&D expenditures, and openness to international trade all had statistically significant impacts on productivity; (ii) regulations were more effective if they were enforced more strictly, a seemingly obvious finding emphasised by Boehm and Oberfeld (2020) based on their study of manufacturing plants in India, known for its weak enforcement practices; and (iii) interactions reinforced the impact of some pairs of policies, suggesting significant complementarities among policies..

Among the more significant recent studies, Bambalaite et al. (2020) stressed the importance of relaxing entry restrictions into professional and service occupations and showed that relaxation enhances aggregate productivity growth through two channels: it enhances productivity at affected firms, and it induces labour reallocation toward more productive firms. Hermansen (2020) explored the prevalence and adverse effects on job mobility, earnings and productivity of occupational licencing, and non-competition agreements in the US. Demmou and Franco (2020) have examined variation in the quality of the governance of access to infrastructure services and pro-competitive regulation in network industries and found both to generate strong productivity growth in downstream industries, again through two channels: sound governance improves the productivity of firms operating in network industries, and it magnifies the downstream productivity effect. The OECD continues to publish studies on the role of institutions, and the most recent findings are available at http://www.oecd.org/economy/OECD-Economics-Department-Working-Papers-by-year.pdf.

Geography

The influence of geography on economic performance has been acknowledged for a very long time. Adam Smith (1776 [1937]), in Book I, Chapter III wrote of the attractions of great towns and water-carriage along the seacoast of the Mediterranean or on the banks of navigable rivers such as the Nile in Egypt and the Ganges in Bengal. He contended that both geographic features expanded the size of the market that otherwise limits the division of labour. Alfred Marshall (1890 [1961]), in Book IV, Chapter X wrote of the concentration of specialised industries in particular localities, citing the breeding of canaries in a small remote village in western Tyrol. He echoed Smith by noting that the advantages of localisation of industry resulted from physical conditions such as the character of the climate and the soil, or easy access by water, and from advantages of proximity to similar and complementary skilled trades. He inquired if these advantages could be maintained by "…the concentration of large numbers of small businesses of a similar kind…" or if these advantages would spur aggregation into "…a small number of rich and powerful firms…", which in turn led him to distinguish internal from external economies. Smith's and Marshall's association of these locational advantages with productivity gains was implicit rather than explicit, but the latter would eventually follow.

Recent developments are based largely on the "new" economic geography initiated by Paul Krugman (1991a, 1991b), recipient of the 2008 Nobel Prize in Economic Sciences. Later Krugman (1998), in a special issue of *Oxford Review of Economics and Policy* devoted to the new economic geography, provided an insightful look back to the beginnings. He described the new economic geography as a genre of research directed to an investigation into the "…geographical concentration of manufacturing based on the interaction of economies of scale with transportation costs", and he created a "…not entirely imaginary history…" that reflected the observations of Smith and Marshall. Subsequent investigations have been global as well as local.

Sachs et al. (2001) used sophisticated mapping software to create five global climate zones. They found both production and productivity, defined as GNP per capita, to be highly concentrated in the coastal regions of temperate climate zones, with per capita income in these regions over twice the global average. William Nordhaus (2006), co-recipient of the 2018 Nobel Prize in Economic Sciences, narrowed the geographic focus from climate zones to cells, by using mapping software to create over 15,000 geographic cells, each with area bounded by 1° latitude by 1° longitude contours. He found a strong negative relationship between mean temperature and output per capita, but a strong positive relationship between mean temperature and output per area, a key variable from a geographic and ecological point of view, up to approximately 5 °C, a paradox he labelled the climate-output reversal. Both Sachs et al. and Nordhaus cautioned that geographic features are not the sole determinants of economic performance, which also depend on the social and economic institutions within which production occurs. Porter (1990, 1998, 2000) also entered the fray, with his introduction of "clusters", geographic concentrations of interconnected companies and institutions in a particular field that create a competitive advantage—think of Silicon Valley or the California wine cluster or the Italian leather fashion cluster. Porter argued that clusters offer advantages in efficiency, effectiveness, flexibility, and innovation enhanced by complementarities among the activities of cluster participants, thereby increasing the productivity of all participants.

Two local applications are worthy of note. Andersson and Lööf (2011) used data on all manufacturing firms with ten or more employees in Swedish municipalities during 1997–2004 to relate labour productivity and location, which they defined as a functional region consisting of several municipalities that together form an integrated local labour market. They found significant economies of agglomeration, with firms located in larger regions being more productive, controlling for human and physical capital, firm size, ownership structure, industry classification, and other variables. This productivity advantage was augmented by a dynamic learning effect, with current agglomeration enhancing future productivity. In a cautionary tale, Au and Henderson (2006) noted that institutions can constrain as well as enhance performance, warning that restrictions on migration constrain agglomeration and productivity, citing China, which has severely limited rural–urban migration as an example. The authors found that these restrictions caused many Chinese cities to be undersized, sacrificing large potential gains in labour productivity from increased agglomeration.

Another example of barriers to labour mobility, and hence aggregate productivity, of current concern involves the elasticity of housing supply, usually through strict zoning laws. Hsieh and Moretti (2019) have examined the effects of stringent restrictions to new housing supply in high productivity cities such as San Francisco and San Jose. These restrictions constrain the number of workers with access to high productivity jobs, resulting in labour misallocation and reduced aggregate labour productivity. The authors studied 220 metropolitan areas from 1964 to 2009, and they found that relaxing, but not eliminating, these barriers to labour mobility would increase the growth rate of aggregate output by over a third, increasing US GDP in 2009 by 3.7%. Garcia Marin et al. (2021) used data from China, Brazil, and the US to confirm that lifting housing supply restrictions raises aggregate total productivity, and to show that it enhances export intensity by allowing exporting firms to locate in large cities, magnifying the productivity gains.

Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies

There exists a distribution of productivities in any group, which reduces aggregate productivity beneath its potential to a degree depending on its variance and its skewness. Both properties of the distribution create a policy challenge.

Productivity dispersion is not a recent phenomenon. Early studies appeared in *Monthly Labor Review*. In addition to the studies of Squires (1917) and Stern (1939) discussed in section "Evidence", Stewart (1922, 1924), who was US Commissioner of Labor Statistics at the time, summarised productivity studies in a wide range of industries, including cotton mills, sawmills, brickyards, blast furnaces, agriculture, and copper mining. In a sample of over 1,000 copper mines, for example, enormous dispersion in labour productivity existed, with the least productive 15% of miners producing 30 pounds per day and the most productive 15% producing 120 pounds per day. Subsequent studies summarised in the *Productivity Measurement Review* also revealed large inter-plant dispersion in labour productivity and financial performance as measured by unit cost, unit labour cost, and return on assets; two of these studies are discussed in section "The Brief Flourishing of the European Productivity Agency". Ingham (1961) summarised the methodology used in these studies and surveyed some of the findings.

Interest in productivity dispersion, its persistence, and its cost continues into the twenty-first century, enhanced by improved methodologies and vastly expanded databases. Syverson (2011) surveyed the literature to 2011 and found evidence of "ubiquitous, large and persistent" productivity dispersion, and the literature has grown rapidly since, quantifying the dispersion, investigating its time path for convergence or divergence, and searching for its sources. Bartelsman and Wolf (2018) discussed subsequent dispersion measures and findings, without decomposing dispersion into within-sector and between-sector contributions, a significant omission since some sectors are widely believed to have higher productivity levels than others, an example being IT-producing and IT-using sectors. However, since productivity can refer to labour productivity or total productivity, and dispersion can be defined asymmetrically with respect to time, as is usual, or symmetrically, as is less common but analytically problematic, a rich methodological literature has developed in the past decade. In his contribution to this *Volume* Balk (in press) provides a succinct summary of the relevant literature and contributes to its continued development in two ways. He derives symmetric decompositions of aggregate output and labour productivity growth, in levels rather than indices that avoid what he describes as "…terms that can be considered as mathematical artefacts, without economic meaning". He then shows how his decomposition of labour productivity growth can provide a convenient foundation for an analysis of dispersion and misallocation, which in turn provides a basis for a search for the drivers and costs of misallocation and appropriate policy remedies. Productivity and its dispersion also can be measured using value functions, as in section "Parametric and Non-parametric Value Functions", and Balk and Zofío (2020b) have developed symmetric decompositions of cost dispersion.

The subsequent literature has followed two occasionally intersecting paths. One path explores misallocation and reallocation, and their effects on productivity. The other explores productivity frontiers and dispersion beneath them. Both strands have found productivity dispersion and resource misallocation to be quantitatively significant, and to exert a proportionate dampening effect on aggregate productivity. I consider the two paths sequentially, the second in more detail than the first.

In an editorial to a special issue of *Review of Economic Dynamics* devoted to misallocation and productivity, Restuccia and Rogerson (2013) provided an overview of the first path, an overview they updated and enriched in Restuccia and Rogerson (2017) and Restuccia (2019).

The essence of this approach is to shift the question from the rhetorical "why are we so rich and they so poor?" to "why are we not richer?" The answer is that misallocated resources reduce our productivity beneath its potential, and reallocation can increase it. The literature devoted to the misallocation component is richer in evidence than that devoted to the reallocation component, perhaps because there is more of it. The simplest, but not the only, way to think about misallocation is via failure to optimise, failure to satisfy first-order conditions for achieving some objective mentioned in section "Analytical Foundations of Productivity Measurement". If the extent of failure in one organisation differs from the extent of failure in another, resources are misallocated, driving aggregate productivity beneath its potential. Efficient reallocation would transfer resources from the organisation exhibiting greater failure to the other organisation, raising aggregate productivity.

Misallocation has many sources, including international trade barriers, product and labour market regulations, credit constraints and other credit market imperfections, restrictions on housing supply, and heterogeneous costs of doing business. Empirical analysis has investigated the magnitudes of the impacts of these misallocations on productivity, which predictably have been negative, ranging from negligible to large. The limited evidence on the effects of reallocation suggests, also predictably, that it raises productivity. Hsieh and Klenow (2009) estimated that reallocation due in part to reforms raised productivity growth by up to 2% in China during 1998–2005, but misallocation worsened in India despite reforms during 1987–1994, reducing productivity growth by about 2%. In a counterfactual exercise, they reduced misallocation in both countries to the level observed in the US and calculated that it would generate huge productivity gains of 30–50% in China and 40–60% in India. Dias et al. (2016) found large potential gains from within-industry reallocation in Portugal leading up to the Eurozone crisis in 2009. They found large and growing misallocation that caused large and growing total productivity dispersion. Between 1996 and 2011, potential GDP loss due to misallocation increased from 48 to 79%, trimming potential GDP growth by 1.3% pa during the period. Calligaris et al. (2018) studied misallocation across the universe of Italian incorporated companies during 1993–2013 and found strong evidence of growing misallocation. Had misallocation remained unchanged over the period, total productivity would have been 18% higher than it was in 2013, which would have translated into a 1% higher annual growth rate. They also found that firms that invested relatively heavily in intangible capital were inefficiently small and under-resourced but keeping up with the technological frontier, while firms with a large share of workers under the Italian wage supplementation scheme or were family-managed or financially constrained were inefficiently large and over-resourced and lagging behind the technological frontier. García-Santana et al. (2020) highlighted cronyism as a source of misallocation in Spain during 1995–2007, in which the most rapid economic growth in at least 150 years occurred despite negative productivity growth and was due solely to factor accumulation. Using a quasi-universe of Spanish firms during the period, they found deteriorating allocative efficiency so severe that, had it remained at initial levels, productivity growth would have been 2.4% pa rather than −0.7% pa. Most interestingly, they found that industries that suffered the largest increases in misallocation were those industries in which connections with public officials were most important for success. They concluded that their findings provide novel evidence on the role of crony capitalism in the economy. Many more such studies exist, and the message rarely varies. Misallocation is ubiquitous, worse in some sectors and in some countries than in others, decreasing in some instances and increasing in others, and it has a potentially large adverse impact on productivity and therefore standards of living.

The second branch uses micro productivity data to construct macro productivity frontiers, and to characterise frontier units and to explore the distribution of units beneath the frontier. Much of the empirical literature uses firm data to create national and global productivity frontiers, defined as the most productive national and global firms in an industry. For example, Andrews et al. (2015) defined national frontiers as the 10 most productive firms within each country, industry, and year, and the global frontier as 50 or 100 most productive firms within each industry and year. In other studies, absolute numbers are replaced by fixed percentages. The two definitions generate very similar productivity gaps between the best and the rest, although they have a common disengagement from the theory of production and its extension to production frontiers. It would be worthwhile therefore to compare empirical findings obtained with these two ad hoc definitions of frontier firms with those obtained using theory-based econometric or mathematical programming frontier techniques summarised in Sickles and Zelenyuk (2019). To facilitate the comparison, Nguyen et al. (in press) survey alternative software packages available for estimating econometric frontiers, with an emphasis on Stata.

Once national and global frontiers have been established, by definition or by estimation, the next step is to characterise national and global frontier firms and explore the productivity gaps between global and national frontier firms and other firms. The eventual challenge is to explain the gaps, frequently in terms of the speed and breadth of technology diffusion, and to design appropriate government policies to enhance diffusion and narrow the gaps. Criscuolo (2015) provided an accessible introduction to the literature, in which she stressed that large and growing productivity gaps are the consequence of disparate abilities to innovate, to combine technological, organisational, and human capital improvements, or as she puts it, "[s]ome firms clearly 'get it' and others don't…" Andrews et al. (2015, 2016) used the OECD-Orbis database to study firms in 23/24 OECD countries at the two-digit level during 2001–2009/1997–2014. They found global firms to be much more productive than other firms, larger but younger, more profitable and more capital- and patent-intensive, and more likely to be part of a multinational group than other firms. Over time, the productivity gap separating frontier firms from other firms has been growing; the gap between global and national frontier firms has been narrowing, but best practices, including the use of advanced technologies, have not diffused beneath national frontier firms. They identified policies designed to narrow both gaps, including the promotion of pro-competition legislation, enhancing education opportunities, introducing R&D tax subsidies, reducing the stringency of employment protection regulation, and revisiting bankruptcy laws that protect extreme laggards (a group I explore below, in which extreme laggards are called "zombies").

Several OECD studies have explored the role of digitalisation in narrowing the productivity gap, which they call the digital divide that instead of narrowing has been widening in the last decade. Berlingieri et al. (2017), in a study using firm-level data across 16 OECD countries during the first decade of the twenty-first century, found digitalisation to be one of the drivers of two divergences, those of productivity and wages, with most of the increased dispersion in both being within rather than across sectors. Significantly, they found the increased dispersion in productivity to be strongly correlated with increased dispersion in wages. Andrews et al. (2018) identified the primary drivers of digitalisation. Using a broad sample of 25 industries in 25 European countries during 2010–2016, they found that business capabilities, such as managerial and technical skills in combination with high performance work practices, and the institutional provision of incentives, such as access to financing and low barriers to entry and exit that contribute to a competitive business environment, both promote digital adoption, and they found strong complementarities between the two. Gal et al. (2019) characterised the productivity gains from digital adoption. They found that adoption of five digital technologies in an industry created substantial productivity gains in firms, but complementarities between digital technologies and skilled labour, organisational capital and other intangibles meant that most productivity gains accrued to already productive firms, thereby widening the digital divide both within and between industries. Sorbe et al. (2019) surveyed a range of policies designed to increase digitalisation, responsible for approximately half of the digital divide. Among the policies they proposed were (i) reform of telecommunication sectors to enable cheaper access to high-speed internet; (ii) increased training, of management and low-skilled employees; (iii) reduced entry costs, including financing constraints; (iv) enhanced efficiency of insolvency regimes; and (v) leading by example, by improving digitalisation of government to exploit synergies between public and private sectors. They also echoed previous writers by stressing the strong complementarities among policies.

Moving away from advanced OECD economies, Maue et al. (2020) took the microeconomic approach to an extreme by studying productivity dispersion and persistence among nearly 10,000 agricultural plots on over 12,000 smallholder farms across four countries in Africa. They used a conventional log-linear production function approach outlined in section "Parametric Production Functions", in which the residual provides a measure of productivity. Their initial survey-based finding was of much larger productivity dispersion, and much lower persistence, than typically found in non-agricultural sectors. However, when they used surveys, satellite information, and other sources to correct for initial measurement error, primarily in output, they found substantial measurement error, correction for which reduced initial estimates of dispersion by roughly half, and roughly doubled initial estimates of persistence, putting estimates of both in line with stylised facts of non-agricultural productivity growth in developing countries.

Most studies have found the distribution of firm productivities to be skewed as well as dispersed, frequently with a long left tail. Whenever firm productivities are negatively skewed, firm employment is likely to be negatively skewed. The literature has distinguished laggard firms, comprising firms in the least productive 40% of the distribution, from unhealthy firms, dubbed zombie firms, those in the bottom 5% or 10% of the distribution. Zombies have been defined in various ways, usually in terms of one or more measures of profitability; all are financially unwell and extreme productivity laggards. Berlingieri et al. (2020) have studied productivity dispersion among laggard firms, and have warned against equating laggard firms, which tend to be younger and smaller than average, with zombie firms, which tend to be older and larger than average. Laggards present a public policy challenge, but zombies create a more serious challenge; they have become the subject of growing public attention and scholarly research, most emphatically during the 2020 pandemic-induced depression. As an example of growing public attention, *The Economist*, which calls them the "corporate undead", has published literally scores of articles about zombie firms and zombie jobs recently.

Zombies were spotted first in Japan during its lost decade of the 1990s. Caballero et al. (2008), Fukuda and Nakamura (2011), and Goto and Wilbur (2019) all studied the causes and consequences of Japanese zombie firms, and all related a similar story. To meet Basle capital standards, many banks continued to extend credit at subsidised rates to insolvent borrowers, a practice called "evergreening", hoping for eventual recovery of the borrowers or government assistance. Caballero et al. showed that approximately 30% of borrowers were kept alive by this practice by the turn of the century. Consequently zombie-infested markets became congested, with less productive zombies occupying market share that discouraged entry and investment by more productive firms and lowering aggregate productivity and slowing Japan's eventual recovery. A natural question to ask is what happened to the zombies after the turn of the century. Fukuda and Nakamura, and Goto and Wilbur, showed that their eventual bankruptcy was rare, and that most recovered satisfactorily. They attributed this outcome to corporate restructuring, which involved shrinking zombies through reducing employee numbers, mainly through dismissals, and shedding fixed assets.

Zombies were discovered next in Europe. Against a backdrop of declining productivity growth in OECD countries, Andrews et al. (2017, Andrews & Petroulakis 2019), Adalet McGowan et al. (2017, 2018), and Gouveia and Osterhold (2018) used firm-level panel data to show that the productivity gap between frontier firms and laggard firms has varied across countries, and has grown rapidly in the twenty-first century, with zombie firms occupying a growing market share in all countries and constraining the growth of more productive firms. The adverse impact on healthy firms has worked through two channels, by reducing productivity-enhancing capital reallocation, and by crowding out credit, both of which reduce aggregate productivity beneath its potential. Banerjee and Hofmann (2018, 2020) studied zombies in 14 advanced economies from the 1980s to 2016. The share of zombies among all firms increased from around 2% in the late 1980s to 15% in 2017, the probability of remaining a zombie from one year to the next increased from 60 to 85%, and their prevalence increased during the 2008 financial crisis. Roughly one quarter of zombies exited and ceased operating. Recovering zombies remained weak, at high risk of relapse, and significantly less productive than other incumbent firms. The authors attributed the rise and survival of zombies to weak banks that rolled over loans rather than calling them or writing them off, in conjunction with low interest rates that reduced borrowing costs and lowered the pressure on zombies to restructure or exit. They found zombies to be less productive than other incumbent firms, thereby reducing aggregate productivity beneath its potential by crowding out investment and employment in healthy firms and potential entrants.

Storz et al. (2017) and Schivardi et al. (2017) studied zombies in euro area periphery countries, including the southern tier countries of Portugal, Spain, Italy, and Greece, as well as Ireland and Slovenia. Storz et al. focused on bank stress as a determinant of the prevalence of zombies, with bank stress a function of capitalisation, non-performing loan ratios and return on assets and other financial indicators. They found variation in bank stress negatively correlated with financial health of zombie borrowers, a result that was robust to alternative definitions of bank stress and zombie financial health. In their study of Italian zombies during the financial crisis, Schivardi et al. found stressed banks were significantly more likely than healthy banks to continue lending to zombies, extending their survival and increasing aggregate productivity dispersion, although the adverse effect of the credit misallocation on the growth of healthy firms was modest. Osterhold and Gouveia (2020) studied zombie firms in Portugal, a valuable case study because Portugal has reduced exit barriers more than most OECD countries. They found a large presence of zombie firms, which were less productive than other firms and depressed aggregate productivity by stifling resource reallocation. They also found that the reduction in exit barriers enhanced exit of the least viable zombie firms and encouraged restructuring of the remaining zombie firms. They also raised an important public policy issue, as zombie firms accounted for over 20% of employment and capital in some sectors, creating a need to design a policy mix that minimises the social costs of exit.

Weak and impaired banks have played a crucial role in these studies, with zombie firms more likely than healthy firms to be linked to weak banks that have exercised forbearance by evergreening loans to zombie firms to avoid realising losses on their balance sheets. Variation in the design of insolvency regimes (e.g., the treatment of non-performing loans and failed entrepreneurs) also has influenced the survival of zombie firms, by enhancing or impeding corporate restructuring and liquidation and by improving creditor recovery rates. This has widened productivity gaps and created differential barriers to restructuring and entry of productive firms and exit of zombie firms. Reform of insolvency regimes along several dimensions in some countries has spurred productivity-enhancing creative destruction and created job displacement, a new twist on the old machinery question, although reform also has created higher non-zombie employment growth and an improved matching of skills with jobs, and these effects have varied across countries as well. However, there remains much scope for improving the design of insolvency regimes, through actions to reduce bank forbearance of non-performing loans and to enhance retraining and job search. Keuschnigg and Kogler (2020) succinctly described the ability of strong banks to fulfil their Schumpeterian role by allocating credit to productive uses and away from zombies, and the inability of weak banks to do either. They provided European evidence of the crucial roles national banking systems have played in both promoting and deterring the reallocation of credit and their influence on aggregate productivity growth.

Zombies have not gone away, and much subsequent public attention and scholarly research have examined the role of the financial systems and insolvency regimes in fulfilling their Schumpeterian role of creative destruction by facilitating the exit or restructuring of zombie firms and zombie jobs, and the policy options available to improve the performance of financial systems and insolvency regimes. The large-scale government and central bank interventions in response to the pandemic depression discussed in section “Productivity and the Pandemic Depression” have magnified concerns about, and coverage of, the growth of zombie firms and their depressing effect on aggregate productivity.

Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth

Nearly a century ago Fagan, C. Clark, and other writers cited in section "Social Concerns" favoured the incorporation of indicators of non-market goods such as leisure and natural resource depletion into a measure of a country's social output. Even earlier, as Sandmo (2015) has noted, the Marquis de Condorcet, Parson Malthus, and several other economists from the eighteenth and nineteenth centuries wrote about natural resources and their growing scarcity, although apparently none of these early writers proposed incorporating resource depletion, pollution, and other environmental outcomes into a holistic measure of a country's social output, much less its productivity. In his "biography" of the subject, Hulten (2001; 33–35) referred to the boundary of productivity analysis, noting that these variables extend "…far beyond the boundaries of the market economy…" and incorporating them would be "…an impossibly large order to fill". Hulten was referring specifically to the then-recent green GDP proposal of Nordhaus and Kokkelenberg (1999) to expand the national accounts to include environmental indicators, although his argument applies to the incorporation of leisure and other non-market goods as well. This section is directed to recent efforts to expand the boundary of productivity analysis beyond the market economy.

Nordhaus and Kokkelenberg wrote of various aspects of, and efforts to implement, environmental accounting. They began by discussing the benefits of augmenting existing economic accounts embodied in the national income and product accounts with satellite natural resource and environmental accounts. Among the benefits they cited were (i) valuable information on the interaction between the environment and the economy; (ii) information showing whether stocks of natural resources and environmental assets were being used in a sustainable manner; and (iii) information to guide regulatory and tax policies. Satellite accounts would include accounts for subsoil mineral assets such as coal, petroleum, and gas, accounts for renewable resources such as timber, and environmental resources such as clean air and water. For some of these variables, "…novel valuation techniques…" would be required for their inclusion in the satellite accounts.

The paucity of valuation techniques mentioned by Nordhaus and Kokkelenberg is a long-standing concern in environmental research involving quantities as well as prices. Significant scientific progress has occurred in the measurement of quantities of all three types of resources, but measurement of prices has proved to be a significant challenge, for natural and environmental resources and other non-market goods. However, for purposes of environmental productivity analysis, the concern can be allayed somewhat. Incorporating non-market variables such as greenhouse gas emissions into an index number approach to productivity analysis summarised in section "Methods" does require prices with which to weight quantity changes. However, incorporating non-market variables into an analytical approach to productivity analysis summarised in section "Non-parametric Distance Functions" does not require price weights, and this approach has become a popular method for conducting environmental productivity analysis. Coelli et al. (2007), Dakpo et al. (2016), and Førsund (2018) provided comprehensive overviews of the modelling issues involved, the essence of which is easy to summarise. Two technologies exist, one for the production of desirable outputs, and the other for the generation of undesirable by-products, and the two technologies are linked in two ways, by shared materials inputs (aka natural capital; see the discussion below) and by the need to satisfy the materials balance condition.

Interest in incorporating environmental variables into augmented national accounts, and then into an environmental productivity analysis, has mushroomed since Hulten wrote of the boundary of productivity analysis.

Costanza et al. (2009) echoed the calls of earlier writers in section "Social Concerns", quoting Kuznets extensively, for the development of improved indicators of progress and well-being that extend beyond GDP. They then surveyed a compendium of related studies of three types: (i) indices that adjust GDP such as the Genuine Progress Indicator; (ii) indices that exclude GDP such as the Ecological Footprint; and (iii) indices that include GDP such as the Human Development Index. Simultaneously, and to much greater acclaim, perhaps because the lead author was a co-recipient of the 2001 Nobel Prize in Economic Sciences, Joseph Stiglitz et al. (2009), in what has become known as "The Stiglitz Report", wrote of the limits of GDP as an indicator of economic performance and social progress. They listed a litany of omissions, most of which C. Clark had mentioned in 1940, including leisure time, inequality of income, wealth, and opportunity, depletion of exhaustible resources and environmental degradation, and trust and social capital. The title of their report notwithstanding, although they stressed the *need* for measurement, they did not suggest *how* to measure these missing indicators. Perhaps because of this omission, soon thereafter Social Indicators Research (2011) devoted most of a special issue to critiques of the Stiglitz Report, most of which centred on the measurement issue.

Nearly a decade after the Stiglitz Report, Stiglitz et al. (2018) released a sequel devoted to the measurement of what counts for economic and social performance. It placed dual emphases on making better use of available statistics and building foundations for new and improved statistics in areas not adequately covered by official statistics, both with an ultimate objective of improving policy analysis and decisions. Prominent among the areas were the several dimensions of inequality, environmental vulnerability and resilience and the sustainability of growth, and social capital and trust, and within each area, it evaluated the quality of available metrics. The report concluded with twelve recommendations on the way ahead for measuring well-being; all but two stressed the need for improving the suite of metrics to be included on the dashboard of indicators with which to guide policy. A companion volume, OECD (2018a) contains essays assessing the adequacy of available metrics and ongoing data challenges for nine areas covered by Stiglitz et al. (2018). To the extent that these metrics become available across OECD countries and through time, they will support a welcome productivity analysis of social economic progress.

The next step has been to progress from an analysis of the adequacy of available metrics to their application to the motivating issue, the measurement of inclusive green growth, an imprecise concept sufficiently malleable to suit a range of policy objectives. The Green Growth Knowledge Platform (2016) has defined it as a multi-faceted concept combining economic growth, environmental sustainability, and social inclusiveness. The OECD (2014) has interpreted the inclusive component as allowing individuals to contribute to economic growth and to receive equitable benefits from it, avoiding inequality of income, wealth, opportunity, and health outcomes. The OECD Green Growth Indicators (2017) defined the green component of inclusive green growth as progress toward four objectives, establishing a low carbon, resource-efficient economy, maintaining the natural asset base, improving people’s quality of life, and implementing appropriate policies toward meeting the first three objectives. It measured progress with 41 Main Indicators, 27 of which are available for most OECD countries.

Setting inclusiveness aside for the moment, it seems appropriate, and tractable, to consider green growth as economic growth constrained by regulations protecting the environment and policies favouring investment in green technologies, the primary channel being the productivity growth component of economic growth. One policy issue then becomes whether stringent environmental regulations undermine or enhance productivity growth, essentially a renewal of the much-maligned Porter (1991) hypothesis. Ambec et al. (2013) provided a 20th anniversary survey of the extant empirical evidence on the hypothesis, which has been decidedly mixed. Support for a weak version, which states that environmental regulation stimulates environmental innovation (which may or may not improve financial performance or raise productivity) has been strong. Support for a strong version, which states that properly designed environmental regulation induces cost-reducing innovation that more than compensates for the cost of compliance and improves the financial (and presumably the productivity) performance of firms, has proved elusive.

More recent OECD evidence suggests a nuanced intermediate outcome, that properly designed environmental regulations can benefit the environment without financial sacrifice or loss in productivity. The OECD has developed a composite index of environmental policy stringency (EPS), details of which are available at http://oe.cd/eps. Albrizio et al. (2017) have used this index to test the strong version of the Porter hypothesis. With a sample of 17 OECD countries and 10 manufacturing sectors during 1999–2009, they found a tightening of environmental policy stringency to have had a positive effect on industry-level productivity growth in countries where an industry was close to the global productivity frontier, with the effect diminishing with distance to the frontier. At the firm level, they found only one-fifth of firms to have increased productivity, and half of firms to have encountered productivity declines, following a tightening of environmental policy stringency. They suggested that the discrepancy between industry- and firm-level findings may reflect exit dynamics or offshoring. Dechezleprêtre et al. (2020) have pursued the exit dynamics issue; they used the EPS index to examine the effects of energy prices and environmental policy stringency on employment in the OECD during 2000–2014. They found the joint effects to be negative and statistically significant, but small in magnitude, as exit of some firms, which reduced employment initially, encouraged employment growth in surviving firms. The OECD also has developed an index of Design and Evaluation of Environmental Policies (DEEP) to augment its long-standing indicators of product market regulation. Berestycki and Dechezleprêtre (2020) have used this index to measure the influence of energy prices and environmental regulatory stringency on competition, which can occur through two channels: first, by distorting market competition through a differential impact across firms, and second by imposing transaction and administrative costs on all firms that can raise barriers to entry. The authors calculated DEEP for 29 OECD countries in 2018 (Korea is best, Italy is worst), but the OECD has not reported empirical evidence obtained from combining DEEP with EPS to identify channels through which increases in the stringency of environmental policy impact productivity growth. The OECD work on policy stringency has been updated and summarised in OECD (2021).

Cárdenas Rodríguez et al. (2018) have developed a growth accounting methodology that can be employed in conjunction with EPS and DEEP. Starting from a transformation function $H(Y, R, L, K, N, t) \geqq 1$, in which Υ is GDP, R is an undesirable environmental by-product, L, K, and N are labour, produced capital, and natural capital, respectively, and t is a time index, they derive

$$\frac{\partial \ln Y}{\partial t} - \varepsilon_{YR}\frac{\partial \ln R}{\partial t} = \varepsilon_{YL}\frac{\partial \ln L}{\partial t} + \varepsilon_{YK}\frac{\partial \ln K}{\partial t} + \varepsilon_{YN}\frac{\partial \ln N}{\partial t} + \frac{\partial \text{lnEAMFP}}{\partial t}.$$

The left side is pollution-adjusted GDP growth, consisting of GDP growth less an adjustment for pollution abatement growth. The right side is input growth, consisting of growth in the contributions of labour, produced capital, and natural capital, plus environmentally adjusted productivity growth. The latter component has a conventional interpretation in an unconventional setting, as a residual representing that part of pollution-adjusted GDP growth that cannot be explained by growth in the use of the three inputs. As with all residuals, it may contain other unincorporated sources of variation in environmentally adjusted productivity growth. An initial empirical study covered 51 countries over the period 1990–2013. Empirical findings varied widely across countries, with environmentally adjusted productivity growth rates ranging from over 3% pa in Estonia, Ireland, and Lithuania to barely positive in Greece and Turkey. Adjustments to GDP growth for pollution abatement ranged from over 0.5% in the Czech Republic and Germany to less than −1% in Turkey, indicating that the first two countries sacrificed potential economic growth with industrial restructuring that reduced emissions, while Turkey relied on emissions-intensive industries to generate growth.

A third finding measured the growth contribution of natural capital to pollution-adjusted GDP growth, with the Russian Federation and Saudi Arabia most reliant at over 0.5% pa and the UK, Hungary, and Denmark reducing their reliance on natural capital.

The incorporation of natural capital in a production relationship by Cárdenas Rodríguez et al. is not new, but recently renewed interest in it is a valuable analytical and empirical development of growing policy interest. In the 1930s and 1940s, Fagan, C. Clark, and H. S. Davis all wrote of the avoidance of a wasteful use of natural resources, and Spengler identified wasting of the heritage of non-reproducible natural resources as distinguishing (colourfully characterised) sound from unsound growth. More recently Nordhaus, Stiglitz, and colleagues have introduced various aspects of the environment as driving a wedge between economic and social progress; see the discussions in sections "Social Concerns" and "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth". Current interest in natural capital has shifted, from a largely normative view of environmental waste and degradation as a by-product of economic growth to a strictly positive effort to incorporate natural capital as an input into an analytical production relationship capable of empirical estimation and testing. This interest has received a boost from the availability of relevant data from the OECD Green Growth Indicators and data developed by other international agencies.

Ecologists have a broad societal perspective on natural capital. Thus Guerry et al. (2015) have interpreted natural capital and the ecosystem services it provides as society's "life support systems" and have explored the role of natural capital in sustaining human well-being. They advocated the incorporation of natural capital and its ecosystem services into management practices and public policies. Costanza et al. (1997) valued the world's ecosystem services and natural capital at 16–54 trillion USD, nearly all of which is outside the market economy. To put this estimate in perspective, and noting the 23-year difference, the 2020 EU GDP was approximately 15 trillion USD. Kareiva et al. (2015) have proposed a greater incorporation of environmental costs and benefits accruing to and arising from natural capital into corporate accounting and reporting. Economists, on the other hand, tend to take a narrower view of natural capital, interpreting it as they do management, as a missing or hard-to-measure input to production processes. Somewhere in between, scientists, call them environmental economists, seek a balance between the preservation of natural capital and the ecosystem services it provides and the use of natural capital in the production of goods and services valued by society. Common to all views are its measurement and the valuation of its contribution to social and private productivity.

Brandt et al. (2017) and Freeman et al. (2021) have treated natural capital as a (previously) missing input, the omission of which causes something akin to management bias analysed by Hoch and others and discussed in section "Organisation". They used OECD data incorporating the subsoil assets component of natural capital to estimate production relationships. Brandt et al. created an unbalanced panel of countries to compare estimates of productivity *growth* incorporating and omitting natural capital over the period 1986–2008. They found national differences ranging from −0.16 in Norway to +0.21 in Chile and +0.28 in Russia. Negative estimates indicate that natural capital grew faster than the traditional inputs of labour and produced capital, and positive estimates indicate the opposite. Thus, traditional productivity growth overstated inclusive productivity growth in Norway, where natural capital grew three times faster than traditional inputs, and understated it in Russia for the opposite reason. Freeman et al. decomposed natural capital into several components (e.g., coal, gas, oil, etc.) and created a 2011 cross-section of countries to compare productivity *levels*. They found productivity levels including natural capital to fall short of those excluding natural capital for each country among the eleven most resource-intensive countries. Thus, traditional productivity levels in these resource-intensive countries overstate inclusive productivity levels. They found the opposite for all countries having a share of natural capital rents less than that of the US, the reference country.

Productivity in the Cárdenas Rodríguez et al., Brandt et al., and Freeman et al. studies is defined using real GDP as the output, with the input vector including and excluding natural capital. The logical next step is to broaden the output concept to a more inclusive measure of social output advocated by writers ranging from Fagan and C. Clark to Stiglitz and colleagues. Nonetheless, these and other findings provide ample evidence in support of a central message of modern productivity analysis I emphasised in section "Introduction", that its purview has widened greatly through its development from its early narrow focus on some measure of growth in output per capita.

The Future: Confronting Two Challenges of Transcendent Significance

We have gathered data, we have developed techniques, and we have gained experience from applying techniques to data in a rich variety of settings over at least the past century. But we have not encountered a pair of circumstances quite like those of the 2020s, a depression brought on by the COVID-19 pandemic and climate change brought on largely by reliance on fossil fuels to power past economic growth. The World Bank Group (2020a) has characterised the pandemic depression as a "crisis", and Tol (2009) has characterised climate change as "the mother of all externalities". These challenges raise the question of whether they influence productivity adversely or positively, and whether productivity can influence either challenge.

In its October 2020 *World Economic Outlook*, the International Monetary Fund (2020) predicted the pandemic would create a depression with "a long and difficult ascent". It forecast a decline in global economic growth to −4.4% in 2020, with a rebound to 5.2% in 2021. It acknowledged that the depth of the contraction would depend on several unknowns, including the length of the pandemic and resulting lockdowns; the impact of social distancing on spending; the ability of displaced workers to secure employment; the extent of scarring from firm closures and extended periods of unemployment; the introduction of strengthened workplace safety measures that incur business costs; and the impact of reconfigurations of global supply chains on business productivity. In its December 2020 *Preliminary Economic Outlook*, the OECD (2020c) forecast a similar decline in global real GDP to −4.2% in 2020 and an increase to 4.2% in 2021. Like the IMF, the OECD acknowledged substantial uncertainty around its projections. Both the IMF and the OECD stressed the desirability of public policies promoting green investment that would address the two challenges of growing out of the depression and mitigating climate change.

I base the rest of the discussion on the academic and business literatures, with an objective of uncovering conjectures or evidence that may contribute to our understanding of the separate and combined impacts of the pandemic and climate change on productivity. In section "Productivity and the Pandemic Depression", I consider the potential impacts of the pandemic and its associated depression on productivity. Much of the literature on this topic exploits what we have learned from past diseases and from past depressions. In section "Productivity and Climate Change", I consider the potential impacts of climate change on productivity, an event that has no precursor from which to learn. The two phenomena are related; greenhouse gas emissions that contribute to global warming have fallen during the pandemic-induced depression as business activity has declined. Accordingly, in section "Linkages Between the Two Challenges", I consider the potential impacts on productivity of complementarities between the two.

Productivity and the Pandemic Depression

The pandemic depression has renewed interest in past depressions, from the 1918 influenza pandemic through the Great Depression of the 1920s and 1930s to the financial crisis of 2008.

The 1918 influenza epidemic caused tens of millions of deaths, 2% of the global population during 1918–1920, and led to declines in output per capita on the order of 7% (Barro et al. 2020, Beach et al. in press), marginally higher than IMF and OECD predictions for the current pandemic depression. Arthi and Parman (2021) surveyed studies of the long-run economic impacts on health, labour, and human capital of the influenza pandemic, and found substantial scarring effects, including reductions in educational attainment and wages. These and many other studies tend to agree on the measurable economic impacts of the pandemic, and on the lack of evidence concerning its impacts on productivity.

Unlike the 1918 economic downturn, the Great Depression of 1929 was not brought on by a public health crisis. Nonetheless, it has relevance for the 2020 pandemic depression. An enormous amount of research has investigated its economic impacts, from Galbraith's (1955) *The Great Crash 1929* to current working papers. Findings of steep declines in output and employment are common, and these declines have implications for productivity trends, the consensus view being that productivity was pro-cyclical. Ohanian (2001), for example, found an 18% decline in total productivity from 1929 to 1933, less than half of which he attributed to popular factors such as changes in capacity utilisation, changes in the composition of production, reallocation of labour, labour hoarding, and increasing returns. He then conjectured that a decline in production efficiency attributable to decreases in organisational capital, "…the knowledge and know-how firms use to organize production…" such as that surveyed in section "Organisation", may have been responsible for much of the remaining decline. Field (2003) argued that, despite the 1929–1933 productivity decline, the decade from 1929 to 1941 was "*…the most technologically progressive of any comparable period in US economic history*" (italics in the original). He attributed a small part of this achievement to a build-up for World War II, and most of it to an unusually large number of economically significant technological and organisational advances that occurred following the depression. Field's bold assertion has received support from Gordon's (2016) massive study of US growth, although Gordon argued that total productivity growth was even faster in the following decade. Magnitudes aside, these and many subsequent studies have found that productivity was pro-cyclical during the Great Depression.

The prevailing productivity story emanating from the 2008 financial crisis changed from one of pro-cyclicality to one of a-cyclicality. Fernald (2015) found strong growth in US total and labour productivity from 1995 through 2003, followed by much weaker growth beginning in 2003 and lasting through 2013. Thus, slower productivity growth preceded and followed, rather than coincided with, the financial crisis. Fernald attributed the pre-crisis productivity slowdown to the waning of the rapid pace of IT investment and complementary innovations such as business reorganisation that boosted productivity growth beginning in the mid-1990s. Fernald and Wang (2016) added to the sources of a-cyclicality more flexible labour markets that have reduced the need to adjust capacity utilisation, a decline in pro-cyclical reallocations within and across production units, a shift in the structure of the economy from manufacturing to services, and the growing importance of intangible investments in R&D, IT, and other hard-to-measure outputs. Galí and van Rens (2021) also claimed the pro-cyclicality of productivity in the US vanished, and its disappearance was driven by increased labour market flexibility resulting from innovations in job search technology and improvements in information about the quality of job matches. This reduced hiring and firing frictions, allowing firms to adjust employment in response to shocks.

Looking forward with the benefit of hindsight gained from analysing previous economic downturns, a growing number of studies of the pandemic depression have appeared. Many have explored the relationship between health and economic outcomes of the pandemic. Others have examined the impact of the pandemic on output and employment, and the ability of fiscal and monetary policy to minimise the adverse impacts. Still others have studied the impacts on international trade and global value chains. Dieppe (2020) has been one of few to study the impact of the pandemic on productivity. He found a largely negative impact to date, caused by uncertainty that weakened domestic and foreign direct investment, mobility restrictions that slowed the reallocation of labour toward higher productivity employment, and weakened corporate and public sector balance sheets that constrained investment and exacerbated employment losses. Looking ahead, he envisioned productivity-enhancing opportunities for businesses and countries that adopt complementary policies toward the integration of new technologies that automate production, the improvement of human resource management, and the development of financial institutions.

Implicit in the discussion above is a much-debated trade-off between the health and economic outcomes of the pandemic, and the ability of public policies to influence the trade-off. Tisdell (2020), Kaplan et al. (2020), and Acemoglu et al. (in press) have created models of the trade-off with great potential value for productivity analysis in general, and specifically for the analysis of productivity dispersion, productivity gaps, distance to the frontier, the identification of zombies, and even the measurement of holistic productivity change incorporating health outcomes. The models differ in their definitions of health and economic outcomes and in other details, but have a common analytical structure. Geometrically, measure health outcomes such as number of COVID-19 cases or deaths per capita on one axis and economic outcomes such as GDP per capita on the other axis, and introduce cross-sectional or panel data on the two outcomes from countries or regions within a country. The data form a pandemic possibility set consisting of all feasible combinations of the two outcomes, with the set bounded by a pandemic possibility frontier that describes the trade-offs between the two outcomes. Conventional frontier estimation techniques project each country to different points on the frontier, reflecting variation in public policies. Some countries seek to avoid adverse health outcomes by imposing social controls such as restrictive lockdowns, perhaps targeted at certain susceptible groups, and promoting vaccinations, while others seek to avoid economic damage at the cost of adverse health outcomes with generous business and employment stimulus packages. Kaplan et al. and Acemoglu et al. stressed the advantages of targeting, by occupation or age or pre-existing co-morbidities, a strategy Tisdell questioned on freedom of choice and ethical grounds, citing Adam Smith and George Orwell. Independently of social preferences, the discovery and dissemination of new medical technologies have the potential to shift the frontier in a favourable direction. A current example of new technologies that shift the frontier is the application of genetics to medicine, in particular the development of messenger ribonucleic acid (mRNA) vaccines to combat the virus. There is guarded optimism that mRNA may be useful in combatting other conditions, including HIV, rabies, and even cancer. The *Economist* (27 March 2021) surveyed the development, the current significance, and the future potential of these new biomedical technologies.

The Economist (8 December 2020) has speculated that the pandemic might spawn a new era of rapid productivity growth. Their reasoning began with Solow's (1987) celebrated quip that "[y]ou can see the computer age everywhere but in the productivity statistics" and continued with David's (1990) reminder from the economic history community that it takes time for general-purpose technologies to bear fruit. It almost concluded with work of Brynjolfsson et al. (2019, 2021) that built on the contributions of Solow and David. These authors argued that AI is a general-purpose technology, which enabled them to exploit the literature on general-purpose technologies, including the time-frame insights of David, to address a current version of Solow's productivity paradox. The adoption of general-purpose technologies such as AI requires investment in complementary intangibles such as R&D, organisational capital, and workforce training, which tend not to appear in company balance sheets or in national accounts, and this has important consequences for productivity measurement. The role played by intangibles generates what the authors called a *productivity J-curve*. Soon after the adoption of a general-purpose technology, true productivity growth is under-estimated because measured inputs are used to accumulate unmeasured intangibles. Eventually true productivity growth is over-estimated because the unmeasured intangibles generate measurable outputs. The productivity J-curve declines and then increases, measuring the deviation between estimated and true productivity growth.

Brynjolfsson et al. developed the productivity J-curve prior to the onset of the pandemic depression. The contribution of the *Economist* was to extend the idea to the pandemic depression, arguing that the pandemic, despite its economic damage, has quickened the adoption of new technology and made a productivity boom more likely to develop. It cited investment in digitisation and automation, and related adoption of AI, 3D/4D printing, robotic process automation (RPA), and numerous other tangibles and intangibles. If economies remain in the downward-sloping portion of the productivity J-curve, the pandemic has brought the upward-sloping portion forward.

Refocusing from national economies to individual businesses, the pandemic forced companies to adopt work from home, or remote work. The practice spread quickly and broadly, and became one of the most studied features of the pandemic. Remote work reduces commuting time, offers flexible working hours, and may improve work-life balance, each of which may influence productivity. The findings have been decidedly mixed.

McKinsey & Company (2020e, 2020f) summarised the findings of a large international survey, in which nearly half of employees working from home reported higher productivity at home than at the office, although fully one-third of respondents reported inadequate internet service, making the investment in digital infrastructure an important policy challenge. They then argued that the pandemic was a tipping point, with business implementing new technologies and operating systems that enhanced the productivity of remote work, but primarily among the well-educated and well-paid minority of the work force. Similar findings were reported by Slack (2020), a corporate messaging firm, based on a survey of 4,700 home workers in six countries. The survey found that flexible working eliminated the money and time cost of commuting, enhanced workers' work-life balance, and increased their productivity. Davis et al. (2021) also postulated that the pandemic would accelerate the widespread adoption of new technologies that increase the productivity of work from home, the key parameter being the elasticity of substitution between market work done at home and market work done at the office. Bloom et al. (2021) provided additional evidence in support of McKinsey and Davis. They used US patent applications to uncover empirical evidence that remote work has induced innovation toward technologies such as remote interactivity that enhance the productivity of remote work. Additional evidence is surely forthcoming.

However, in a pair of member surveys, the Association of Chartered Certified Accountants (2020, nd) found that the most frequently mentioned impact of the pandemic was *reduced* employee productivity, even though most respondents have implemented flexible work strategies and most believed remote work would be a long-lasting pandemic impact. In their study of the switch from office work to remote work by over 10,000 highly skilled employees in a large Asian IT company, Gibbs et al.

(2021) measured productivity by the ratio of an output measure provided by the company to hours worked on a relevant task. They found a significant increase in hours worked, no significant change in measured output, and a productivity decline of about 20%. They suggested that remote work hampers communication, coordination, and collaboration, and the impact on the productivity of highly skilled professionals may differ from that of other workers due to the nature of the IT job requirements.

The academic literature has evinced an almost instant recognition of the economic and public health significance of the 2020 pandemic depression. In addition to a flood of working papers, two new academic journals devoted to vetted real-time economic analysis of the pandemic have appeared, *Covid Economics*, from the Centre for Economic Policy Research, and *The Economics of the Coronavirus Crisis*, from *Intereconomics/Review of European Economic Policy*.

Productivity and Climate Change

In contrast to the pandemic depression, climate change has attracted a multitude of studies directed toward its impact on productivity, perhaps because it has been occurring for centuries, whereas the pandemic depression began in 2020.

In its Fifth Assessment Report, the Intergovernmental Panel on Climate Change (IPCC) (2014) stated that warming of the climate system is "unequivocal", and human influence on the climate system is "clear". The Panel noted variation in vulnerability to climate change across nations, and across regions and sectors within nations, and stressed the need for complementary policies and actions to promote mitigation of and adaptation to the impacts of climate change. It also emphasised the constraints facing implementation of both, and the likely gaps separating achievements from possibilities. The emphasis on gaps brings to mind something similar to Kaplan's pandemic possibility frontier in section "Productivity and the Pandemic Depression", with climate change substituted for pandemic and inspiration drawn from the literature on productivity dispersion and distance to frontier in section "Productivity Dispersion, Productivity Gaps, Distance to Frontier and Zombies". In a subsequent Special Report, the IPCC (2018) asserted, with high confidence, that global warming is likely to reach 1.5 °C above pre-industrial levels between 2030 and 2052 if it continues to increase at the current rate. To provide an idea of the magnitude of the likely cost required to limit global warming to 1.5 °C target by the IPCC date range, van Vuuren et al. (2020) have constructed a meta-model from climate and integrated assessment models to generate an estimate of the cumulative abatement costs of meeting the target. Using a 5% discount rate, their median estimate of the cost is 30 trillion USD, with a 90% confidence interval of 10–100 trillion USD. Recall from above that the 2020 EU GDP was approximately 15 trillion USD.

I consider two sectors, agriculture, in which the impacts are particularly severe, and business, whose managements must adapt. In both cases, farm-level and firm-level impacts and responses aggregate to national outcomes. For an insider's view on the difficulties encountered in attempting to implement a policy agenda for dealing with climate change, I recommend Garnaut (2019), a readable survey of the economic and political issues involved, with a global perspective set against an Australian backdrop.

Because agriculture is particularly sensitive to the vagaries of the weather, and since crop productivity is commonly measured by easily observable yield, crop output per area, it has attracted a large volume of research into the impacts of climate change on agricultural production. Two recent studies illustrate the diversity of issues involved and the importance of developing flexible models of the relationship. Wang et al. (2019) and O'Donnell (2021) provided empirical evidence on the effects of weather and climate change on US agricultural productivity. The two studies used the same economic data, a state-by-year panel of three outputs and four inputs covering 1960–2004 available at https://www.ers.usda.gov/data-products/agricultural-productivity-in-the-us/, although they used these data to measure productivity very differently. Wang et al. added two climate variables, a temperature-humidity index that measures the effects of extreme heat and humidity on livestock production, and an aridity index that measures the effects of rain deficit on crop production. They found productivity to have been sensitive to *long-term trends* and interstate variation in both climate variables, but that both impacts have diminished through time as states have adapted to changing climate conditions. However, *unexpected shocks* of both types have had substantial productivity impacts. O'Donnell added three different weather variables, a pair of heat indicators, and a precipitation indicator. He found the main drivers of productivity change through time to have been improvements in technology and change in scale and mix efficiency, and the main drivers of productivity variation across states to have been variation in production environments (e.g., soil type and terrain) and variation in scale and mix efficiency. In contrast to the findings of Wang et al., O'Donnell found inter-temporal change and interstate variation in weather variables to have exerted a relatively small impact on US agricultural productivity.

Turning to business, *The Economist* (17 September 2020) called climate change "the great disrupter" and described several channels through which climate change has influenced business behaviour, and how government policies have affected the relationship. Among these channels are the disruption of global supply chains, the regulation and deregulation of emissions, (the paucity of) carbon pricing, and a growing incentive to direct process and product innovation away from fossil fuels and toward the use of renewable resources such as the sun and wind. Many of these issues, and more, have appeared in the business press. In one of many articles on the impacts of climate change on business, McKinsey & Company (2020c) stressed the growing risks to business performance posed by climate change, especially its impacts on global supply chains, which are "optimised for efficiency, not resilience", and suggested business strategies for adaptation. Deloitte (2020) conducted a survey of over 1,000 European CFOs, who revealed growing pressure to act from a broad range of stakeholders. Despite the growing pressure, few companies have analysed risks or have governance structures in place and have reacted largely by pursuing short-term cost-saving strategies and setting emissions reduction targets not aligned with the 2015 Paris Agreement. Deloitte does identify potential business opportunities such as improving energy efficiency, creating new products or services that are less energy-intensive, and enhancing the resilience of their supply chains. Each of these strategies has the potential to improve business productivity, holistically if not conventionally defined.

I now turn to the impacts of climate change on aggregate productivity. Heal and Park (2013) used country-level panel data to derive temperature-driven productivity impacts, and they found significant temperature sensitivity of per capita income that varies with a country's position relative to the temperate zone. In hot zones, the impact of an increase in temperature is large and negative, while in cold zones the opposite happens, both with approximately 3–4% productivity change per degree C. They did not explore the trade and migration possibilities created by these geographically opposing effects. Nath (2020) did;

he studied the impact of temperature on sectoral reallocation and aggregate productivity, using firm-level data across a wide range of countries. His estimates showed that extreme heat reduced non-agricultural output per worker, but by less than in agriculture, implying that hot countries could adapt by shifting resources from agriculture to manufacturing. Simulations suggested that this has not happened, since subsistence food requirements dominate comparative advantage. Climate change draws labour into relatively low productivity agriculture rather than drawing it away, with the perverse reallocation effect exerting downward pressure on global GDP. Cruz Alvarez and Rossi-Hansberg (2021) have documented the wide geographic variation in the impacts of global warming and have predicted large productivity and welfare losses in parts of Africa, India, and Latin America, and gains in Siberia, Alaska, and northern Canada. They emphasised that their magnitudes depend crucially on economic adaptation mechanisms, the extent of migration and inter-regional trade, and endogenous local innovation.

It is also possible to incorporate environmental impacts into an inclusive model of productivity growth. The OECD (2018b) used the augmented growth accounting methodology of Cárdenas Rodríguez et al. (2018) to estimate China's environmentally adjusted productivity growth during 2000–2013 at approximately 2.5% pa, with a declining trend reflecting China's growing reliance on natural resources and ecosystem services to fuel economic growth. Li and Ouyang (2020) used an alternative methodology to estimate green productivity growth in 284 Chinese cities during 2004–2015. They started from the premise developed by Acemoglu et al. (2012) that technical progress may be directed to green or brown technologies. They incorporated three components of technical change, indigenous technical change embedded in the stock of knowledge in patents, technology transfers from foreign direct investment, and absorptive capacity, the ability to assimilate and apply new technology to commercial ends. They found green productivity to have trended downward during this phase of the Chinese extensive development model, which promoted rapid energy- and resource-intensive growth that made China the world's largest contributor to global greenhouse gas emissions. They also found indigenous technical change to have had an adverse impact on green productivity growth, since patents tended to protect existing brown technologies, and they found the impact of technology transfers to have been contextual, depending on a city's per capita income among other determinants. Only absorptive capacity had a positive, albeit small, impact on green productivity. They also found that environmental regulation enhanced green productivity growth in an expanded model of technical change, which provided support for the Porter Hypothesis. A new Chinese economic growth model was enshrined toward the end of the study period, promoting slower green growth with an energy- and resource-saving orientation augmented with restrictive environmental protection policies. Growth has indeed slowed, and the energy- and resource-intensity of GDP have declined. A similar green productivity study quantifying the benefits would be welcome, particularly if it captured the disruption of the pandemic recession.

Several academic journals specialise in either climate change or environmental economics, and both publish studies exploring economic aspects of climate change. In addition to a growing number of working papers, one journal is devoted exclusively to the economics of climate change, *Climate Change Economics*, which recently devoted a special issue commemorating William Nordhaus' receipt of the Nobel Prize in Economic Sciences. Another journal devoted to all aspects of climate change, *Oxford Open Climate Change*, launched in late 2020, and *Economic Policy* has devoted a special issue to the economics of climate change.

Linkages Between the Two Challenges

The Lancet (2020) published an editorial about the two "converging" crises of climate change and the pandemic, noting their common causes of human activity and their common consequences for human health, and stressing the oft-reported observation that the poorest and most marginalised people are the most vulnerable. A related report from *The Medical Journal of Australia* (2020) documented the common causes and consequences in Australia, where temperature extremes and bushfires exacerbated the health effects.

Simultaneous occurrence does not imply causality, but the pandemic depression has slowed greenhouse gas emissions, at least temporarily. However, like all previous depressions, the pandemic depression will end, which has motivated a search for other ways to slow or reverse the growth of greenhouse gas emissions in an environment of economic growth. Numerous proposals have appeared, some of them introduced in section "Expanding the Scope of Productivity Analysis Redux: Inclusive Green Growth" under the heading of inclusive green growth. As a general policy-oriented observation, there has developed a widespread agreement among scholars, if not among politicians, that synergies are there waiting to be exploited with the appropriate policies. Programs to boost economies out of the pandemic depression can promote green growth, for example by switching from fossil fuels to climate-friendly renewable energy sources and complementary technologies, and by investing in the greening of buildings and transport, all examples of directed technical change. However, the gap between scientists and policymakers is yawning, and is emphasised in a wide array of studies. Some, such as Gettins (2020) and Gardiner (2020) point to the urgency of addressing both challenges as distinguishing scientists from policymakers, who appear to lack the urgency in combatting climate change they exhibited in fighting the pandemic. Others, including Pearce (2020) and Spratt and Armistead (2020), point to the need for, and the uncertain likelihood of, a green recovery from the pandemic. As this survey has stressed in other contexts, management matters.

Le Quéré et al. (2020) chronicled the reduction in daily CO_2 emissions during the pandemic depression. They calculated a decrease of 17% through April 2020 relative to mean 2019 emissions, and they predicted emissions reductions ranging from 4 to 7% for all of 2020, with large variances depending on government actions and economic incentives. Liu et al. (2020) and Friedlingstein et al. (2020) predicted that global CO_2 emissions would decline by 8.8 and 7%, respectively, throughout 2020. Forster et al. (2020) have taken a longer-term perspective, predicting a negligible impact of the pandemic depression on global greenhouse gas emissions by 2030, depending on the extent to which the recovery tilts toward green stimulus and reduced fossil fuel investments. Thus, the pandemic depression and climate change have been closely related through late 2020, are expected to be modestly correlated in the short term and conditionally correlated in the long term.

Helm (2020) has summarised the short-term environmental impacts of the pandemic depression and has offered a somewhat nuanced look ahead to some possible long-term consequences. He considered two impacts in detail, the possible re-orientation of public fiscal and pricing policies in a green direction, particularly in energy and transport, and the potential for continued de-globalisation and shortening of value chains initiated by the pandemic. Concerning the first impact, he expressed a preference for pricing environmental impacts over fiscal stimulus programs but lamented that pricing of environmental impacts was more popular with economists than with legislators. Regarding the second impact, he noted that the relative decline of domestic production of five widely traded carbon-intensive goods in the EU and US since China's accession to the WTO was largely replaced by coal exports to China. This practice exported carbon emissions from the EU and US to China, thereby increasing emissions through two channels, from shipping and aviation and from relocating production of carbon-intensive goods away from countries with relatively high environmental standards. From the recent reshoring experience during the pandemic, Helm found grounds for optimism, inferring that de-globalisation may reduce total greenhouse gas emissions.

Many other studies have chronicled the short-term environmental impacts of the pandemic depression and estimated the long-term impacts. However, the long-term impacts depend in large part on the public policies enacted in the interim. Hepburn et al. (2020), Engström et al. (2020), and Agrawala et al. (2020) have considered a range of policies and have evaluated the relative merits of green and brown approaches. Hepburn et al. surveyed a large number of central bank and finance ministry officials and other economic experts from G20 countries on the relative merits of 25 recovery policies, using four criteria: speed of implementation, economic multiplier, climate impact potential, and overall desirability. From their responses, the authors identified five policies having high potential on both economic multiplier and climate impact criteria: clean physical infrastructure investment, building efficiency retrofits, investment in education and training to address both immediate unemployment from COVID-19 and structural unemployment from de-carbonisation, natural capital investment for ecosystem resilience and regeneration, and clean R&D investment. Engström et al. introduced another consideration, an economy's public health objectives, and consistent with concerns expressed through several IPCC Assessment Reports, Agrawala et al. added to public health yet another consideration, an economy's social and distributional policy objectives. This led them to propose a vague "just transition" reminiscent of C. Clark's call for a "just" distribution of the fruits of productivity growth in 1940.

This inclusive interpretation aligns with the OECD's *Focus on Green Recovery* website (https://www.oecd.org/coronavirus/en/themes/green-recovery), which contains numerous current policy papers, policy responses, and blogs, all directed toward the importance of developing public policies that would exploit the synergies, by pursuing green growth. I cite two of several policy-oriented documents. The OECD

(2020a) has proposed "building back better" from the two challenges, in which policies directed toward improving well-being and inclusiveness occupy the central position among a circle of economic and environmental policies. The explanation for centrality is persuasive: centrality is crucial to gaining social and political acceptance of economic and environmental policies. Whereas the OECD (2020a) stressed the importance of the *inclusiveness* of the recovery, the OECD (2020b) stressed the *greenery* of the recovery. It proposed six outcome indicators, with particular significance attached to the share of renewable energy in the energy mix and material productivity, the ratio of real GDP to the consumption of domestic raw materials from natural resources. It proposed seven policy indicators intended to enhance a green recovery, including the usual shop-worn tax, subsidy, and carbon pricing schemes, but also an expansion of environmental R&D expenditure. An objective summary of the OECD's pandemic recovery vision would be that productivity growth has a significant role, provided it is inclusive and green.

The previously cited literature is macroeconomic in nature, and businesses respond to macroeconomic policies with management decisions that make it desirable to explore the business literature linking the two challenges. McKinsey & Company (2020a, 2020b, 2020d) has been at the forefront, claiming that business simply cannot afford to ignore the dual challenge, and set two priorities. The first is to decarbonise. The second involves making operations more resilient and more sustainable, by shortening and diversifying value chains, investing in energy-efficient manufacturing, and increasing digitisation of sales and marketing. Addressing both priorities requires investment, and McKinsey notes that, with near-zero interest rates for the near future, there is no better time than the present for such investments, a sentiment shared in much of the business literature. Numerous sources warn, however, of bottlenecks to investment at a scale necessary to pursue green growth, to reach net zero, or to meet the Paris Agreement target.

The academic literature has shown a growing recognition of the significance of the joint impacts of, and the complementary solutions to, the pandemic depression and climate change. In addition to a rapidly growing number of working papers, at least two academic journals have devoted special issues to the joint challenge, *Environmental and Resource Economics* 76:4 (August 2020) and *Oxford Review of Economic Policy* 36, Supplement 1 (2020).

Conclusions

Productivity analysis has come a long way from its origins in observation in the distant past, as I have chronicled in this survey. I have documented advances in data collection to provide evidence, in measurement to quantify evidence, in analytical modelling to incorporate objectives of and constraints to production activities, and in a century of effort to recognise the critical role of management in productivity analysis and to broaden the scope of productivity analysis beyond the market economy.

In the process of writing this survey, I have also highlighted what I believe are three significant developments within the overall growth of the field of productivity analysis that tend to be overlooked in conventional surveys. One is a growing interest of economists in the productivity performance of individual businesses. This has shown up most visibly in efforts to find the missing management input, which has been a prominent component of business school curricula for generations, although we began to incorporate it only in the 1950s. More recently, we have developed sophisticated models of the causes and consequences, both financial and productivity, of variation in the quality of management practices, backed by large data sets. Moreover, the aggregate productivity of nations and industries that has dominated our research is (almost) simply that, an aggregate of the productivities of individual businesses directed by Drucker's managers, helped or hindered along the way by Koopmans' helmsmen. A second development began with Ricardo's machinery question and refuses to fade away. Technology was long viewed as a source of labour displacement, embodied at one stage by the mechanical cotton picker. But now new technology embodied in robotics, artificial intelligence, and machine learning and other advances in information and communications technology, has of necessity raised the profile of human resource departments assigned the task of accommodating it, thereby illustrating the complementarities involved, and has come to be viewed as an admittedly disruptive source of economic and social progress. A third development is a concomitant growing interest in incorporating non-market activities into productivity analysis. This interest attracted prominent economists to express a range of social concerns in the wake of the Great Depression and has re-emerged nearly a century later among growing environmental concerns expressed in the Inclusive Green Growth movement at the aggregate level, and in the CSR and ESG movements at the business level. It is worth noting that the non-market sector is populated by more than just environmentalists and stakeholders. For example, economic activities within the household have attracted the attention of Becker and a host of prominent economists for a long time, although their productivity consequences have been investigated only recently, and they receive only a brief recognition for this survey.

An unfortunate shortcoming has permeated some of the issues I find most interesting, the occasional inability to explicitly incorporate productivity, the result of the data constraint, or a focus on issues of more immediate concern such as health, a drawback that is nonetheless particularly worrisome for a survey of productivity measurement! Nevertheless, I find two causes for muted optimism. First, these issues revolve around resources and outcomes, regardless of whether they are adequately captured in the data under investigation. From resources and outcomes, it is a relatively short step to the ratio of the two, or the distance between the two, and that distance needs to be traversed. Second, provided productivity is properly measured, and with the important caveat of keeping context in mind, productivity improvements contribute positively to addressing any economic challenge. Hopefully, the current objective of analysing the potential of productivity growth to contribute to the solution of the two simultaneous challenges of transcendent significance, will motivate an enlightened subsequent survey of productivity analysis.

I close on a happy note. Eleven recipients of the Nobel Prize in Economic Sciences, awarded over a half century from 1969 through 2018, appear in this survey, attesting to the significance of productivity analysis and measurement.

References

Abramovitz, M. (1956). Resource and output trends in the United States since 1870. *American Economic Review, 46*, 5–23.

Acemoglu, D., Aghion, P., Burstyn, L., & Hemous, D. (2012). The environment and directed technical change. *American Economic Review, 102*, 131–166.

Acemoglu, D., Chernozhukov, V., Werning, I., & Whinston, M. D. (in press). Optimal targeted lockdowns in a multi-group SIR Model. *American Economic Review: Insights, 3*, 487–502.

Acemoglu, D., Johnson, S., & Robinson, J. (2005). Institutions as a fundamental cause of long-run growth, Chapter 6. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1A). North-Holland.

Acemoglu, D., & Restrepo, P. (2018). The Race between man and machine: Implications of technology for growth, factor shares and employment. *American Economic Review, 108*, 1488–1542.

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives, 33*, 3–30.

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy, 128*, 2188–2244.

Adalet McGowan, M., Andews, D., & Millot, V. (2017). *Insolvency regimes, zombie firms and capital reallocation* (OECD Economics Department Working Papers No. 1399). https://doi.org/10.1787/5a16beda-en

Adalet McGowan, M., Andews, D., & Millot, V. (2018). The walking dead? Zombie firms and productivity performance in OECD countries. *Economic Policy, 33*, 685–736.

Adhvaryu, A., Nyshadham, A., & Tamayo, J. A. (in press a). Managerial quality and productivity dynamics. *Review of Economic Studies*.

Adhvaryu, A., Kala, N., & Nyshadham, A. (in press b). Management and shocks to worker productivity. *Journal of Political Economy*.

Agrawala, S., Dussaux, D., & Monti, N. (2020). *What policies for greening the crisis response and economic recovery? Lessons learned from past green stimulus measures and implications for the COVID-19 crisis* (OECD Environment Working Papers No. 164). https://doi.org/10/1787/c50f186f-en

Ahmad, N., & Koh, S.-H. (2011). *Incorporating estimates of household production of non-market services into international comparisons of material well-being* (OECD Statistics Working Papers 2011/07). https://doi.org/10.1787/5kg3h0jgk87-en

Albrizio, S., Kozluk, T., & Zipperer, V. (2017). Environmental policies and productivity growth: Evidence across industries and firms. *Journal of Environmental Economics and Management, 81*, 209–226.

Alexopoulos, M., & Tombe, T. (2012). Management matters. *Journal of Monetary Economics, 59*, 269–285.

Allen, R. C. (2000). Economic structure and agricultural productivity in Europe, 1300–1800. *European Review of Economic History, 3*, 1–25.

Allen, R. C. (2001). The great divergence in European wages and prices from the Middle Ages to the First World War. *Explorations in Economic History, 38*, 411–447.

Ambec, S., Cohen, M. A., Elgie, S., & Lanoie, P. (2013). The Porter hypothesis at 20: Can environmental regulation enhance innovation and competitiveness? *Review of Environmental Economics and Policy, 7*, 2–22.

Andersson, M., & Lööf, H. (2011). Agglomeration and productivity: Evidence from firm-level data. *Annals of Regional Science, 46*, 601–620.

Andrews, D., Adalet McGowan, M., & Millot, V. (2017). *Confronting the zombies: Policies for productivity renewal* (OECD Economic Policy Paper No. 21). https://doi.org/10.1787/f14fd801-en.

Andrews, D., Criscuolo, C., & Gal, P. N. (2015). *Frontier firms, technology diffusion and public policy: Micro evidence from OECD countries* (OECD Productivity Working Papers No. 02). https://doi.org/10.1787/5jrql2q2jj7b-en

Andrews, D., Criscuolo, C., & Gal, P. N. (2016). *The best versus the rest: The global productivity slowdown, divergence across firms and the role of public policy* (OECD Productivity Working Papers No. 05). https://doi.org/10.1787/63629cc9-en

Andrews, D., Nicoletti, G., & Timiliotis, C. (2018). *Digital technology diffusion: A matter of capabilities, incentives or both?* (OECD Economics Department Working Papers No. 1476). https://doi.org/10.1787/7c542c16-en

Andrews, D., & Petroulakis, F. (2019). *Breaking the shackles: Zombie firms, weak banks and depressed restructuring in Europe* (OECD Economics Department Working Papers No. 1433). https://doi.org/10.1787/0815ce0c-en

Aparicio, J., López-Torres, L., & Santin, D. (2018). Economic crisis and public education: A productivity analysis using a Hicks-Moorsteen Index. *Economic Modelling, 71*, 34–44.

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The risk of automation for jobs in OECD countries: A comparative analysis* (OECD Social, Employment and Migration Working Papers No. 189). https://doi.org/10.1787/5jlz9h56dvq

Arthi, V., & Parman, J. (2021). Disease, downturns, and wellbeing: Economic history and the long-run impacts of Covid-19. *Explorations in Economic History, 79*, 101381.

Association of Chartered Certified Accountants. (2020). *COVID-19 global survey: Inside business impacts and responses*. https://www.accaglobal.com/hk/en/professional-insights/global-economics/Covid-19_A-Global-Survey.html

Association of Chartered Certified Accountants. (n.d.). *COVID-19: The road to recovery?* file:///C:/Users/uqclove1/AppData/Local/Temp/Covid-19_RoadRecovery_SlideStyle-1.pdf

Au, C.-C., & Henderson, V. (2006). How migration restrictions limit Agglomeration and productivity in China. *Journal of Development Economics, 80*, 350–388.

Autor, D., Mindell, D. A., & Reynolds, E. B. (2019). *The work of the future: Shaping technology and institutions*. https://workofthefuture.mit.edu/report/work-future

Autor, D., & Salomons, A. (2018). Is automation labor share-displacing? Productivity growth, employment, and the labor share. *Brookings Papers on Economic Activity* (Spring), 1–63.

Balk, B. M. (1995). Axiomatic price index theory: A survey. *International Statistical Review, 63*, 69–93.
Balk, B. M. (2008). *Price and quantity index numbers*. Cambridge University Press.
Balk, B. M. (in press). Symmetric decompositions of aggregate output and labor productivity growth: On levels, (non-) additivity, and misallocation. this *Volume*.
Balk, B. M., Barbero, J., & Zofío, J. L. (2020). A toolbox for calculating and decomposing total factor productivity Indices. *Computers & Operations Research, 115*, 104853.
Balk, B. M., & Zofío, J. L. (2020). Symmetric decompositions of cost variation. *European Journal of Operational Research, 285*, 1189–1198.
Bambalaite, I., Nicoletti, G., & von Rueden, C. (2020). *Occupational entry regulations and their effects on productivity in services: Firm-level evidence* (OECD Economics Department Working Papers No. 1605). https://doi.org/10.1787.c8b88d8b-en
Banerjee, R., & Hofmann, B. (2018, September). The rise of zombie firms: Causes and consequences. *BIS Quarterly Review*, 67–78.
Banerjee, R., & Hofmann, B. (2020). *Corporate zombies: Anatomy and life cycle* (Bank for International Settlements, BIS Working Papers No. 882). https://www.bis.org/publ/work882.htm
Barro, R. J., Ursúa, J. F., & Weng, J. (2020). *The coronavirus and the great influenza pandemic: Lessons from the 'Spanish Flu' for the coronavirus's potential effects on mortality and economic activity* (NBER Working Paper 26866). http://www.nber.org/papers/w26866
Barseghyan, L. (2008). Entry costs and cross-country differences in productivity and output. *Journal of Economic Growth, 13*, 145–167.
Barseghyan, L., & DiCecio, R. (2011). Entry costs, industry structure, and cross-country income and TFP differences. *Journal of Economic Theory, 146*, 1828–1851.
Bartel, A., Ichniowski, C., & Shaw, K. (2007). How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement and worker skills. *Quarterly Journal of Economics, 122*, 1721–1758.
Bartelsman, E. J., & Wolf, Z. (2018). Measuring productivity dispersion, Chapter 18. In E. Grifell-Tatjé, C. A. K. Lovell & R. C. Sickles (Eds.), *The Oxford handbook of productivity analysis*. Oxford University Press.
Baumol, W. J. (1959). *Business behavior, value and growth*. Macmillan.
Baumol, W. J., Panzar, J. C., & Willig, R. D. (1982). *Contestable markets and the theory of industry structure*. Harcourt, Brace & Jovanovich.
Beach, B., Clay, K., & Saavedra, M. H. (in press). The 1918 influenza pandemic and its lessons for Covid-19. *Journal of Economic Literature*.

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

Bender, S., Bloom, N., Card, D., & Van Reenen, J. (2018). Management practices, workforce selection, and productivity. *Journal of Labor Economics, 36*, s371–s409.

Benner, M. J. (2018). Innovation, management practices, and productivity, Chapter 12. In E. Grifell-Tatjé, C. A. K. Lovell & R. C. Sickles (Eds.). *The Oxford handbook of productivity analysis*. Oxford University Press.

Berestycki, C., & Dechezleprêtre, A. (2020). *Assessing the efficiency of environmental policy design and evaluation: Results from a 2018 cross-country survey* (OECD Economics Department Working Papers No. 1611). https://doi.org/10.1787/482f8fbe-en

Berg, M. (2010). *The machinery question and the making of political economy 1815–1848*. Cambridge University Press.

Bergoeing, R., Loayza, N. V., & Piguillem, F. (2016). The whole is greater than the sum of its parts: Complementary reforms to address microeconomic distortions. *World Bank Economic Review, 30*, 268–305.

Berle, A., & Means, G. (1932). *The modern corporation and private property*. World.

Berlingieri, G., Blanchenay, P., & Criscuolo, C. (2017). *The great divergence(s)* (OECD STI Policy Papers 39). https://doi.org/10.1787/953f3853-en

Berlingieri, G., Calligaris, S., Criscuolo, C., & Verlhac, R. (2020). *Laggard firms, technology diffusion and its structural and policy determinants* (OECD STI Policy Papers 86). https://doi.org/10.1787/281bd7a9-en

Bjurek, H. (1996). The Malmquist total factor productivity index. *Scandinavian Journal of Economics, 98*, 303–313.

Black, S. E., & Lynch, L. M. (2001). How to compete: The impact of workplace practices and information technology on productivity. *Review of Economics and Statistics, 83*, 434–445.

Blank, J., & Merkies, A. H. Q. M. (2004). Empirical assessment of the economic behaviour of Dutch general hospitals. *Health Economics, 13*, 265–280.

Bliss, J. H. (1923). *Financial and operating ratios in management*. The Ronald Press Co.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., & Van Reenen, J. (2019). What drives differences in management practices? *American Economic Review, 109*, 1648–1683.

Bloom, N., Davis, S. J., & Zhestkova, Y. (2021). COVID-19 shifted patent applications toward technologies that support working from home. *American Economic Review Papers & Proceedings, 111*, 263–266.

Bloom, N., Genakos, C., Martin, R., & Sadun, R. (2010). Modern management: Good for the environment or just hot air? *Economic Journal, 120*, 551–572.

Bloom, N., Kretschmer, T., & Van Reenen, J. (2009). Work-life balance, management practices and productivity, Chapter 1. In R. B. Freeman & K. L. Shaw (Eds.), *International differences in the business practices and productivity of firms*. University of Chicago Press.
Bloom, N., Sadun, R., & Van Reenen, J. (2012). Americans do IT better: US multinationals and the productivity miracle. *American Economic Review, 102*, 167–201.
Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics, 122*, 1351–1408.
Bloom, N., & Van Reenen, J. (2011). Human resource management and productivity, Chapter 19. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 4B). Elsevier.
Boehm, J., & Oberfeld, E. (2020). Misallocation in the market for inputs: Enforcement and the organization of production. *Quarterly Journal of Economics, 135*, 2007–2058.
Brandt, N., Schreyer, P., & Zipperer, V. (2017). Productivity measurement with natural capital. *Review of Income and Wealth, 63*, s7–s21.
Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics, Chapter 1. In A. Agrawal, J. Gans & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda*. University of Chicago Press.
Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics, 13*, 333–372.
Caballero, R. J., Hoshi, T., & Kashyap, A. K. (2008). Zombie lending and depressed restructuring in Japan. *American Economic Review, 98*, 1943–1977.
Cai, J., & Wang, S.-Y. (2020). *Improving management through worker evaluations: Evidence from auto manufacturing* (NBER Working Paper 27680). http://www.nber.org/papers/w27680
Calligaris, S., Del Gatto, M., Hassan, F., Ottaviano, G. I. P., & Schivardi, F. (2018). The productivity puzzle and misallocation: An Italian perspective. *Economic Policy, 33*, 635–684.
Cárdenas Rodríguez, M., Haščič, I., & Souchier, M. (2018). Environmentally adjusted multifactor productivity: Methodology and empirical results for OECD and G20 countries. *Ecological Economics, 153*, 147–160.
Chandler, A. D. (1962). *Strategy and structure: Chapters on the history of the industrial enterprise*. The MIT Press.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research, 2*, 429–444.
Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1971). Conjugate duality and the transcendental logarithmic function. *Econometrica, 39*(1971), 255–256.

Clague, E., & Couper, W. J. (1931). The readjustment of workers displaced by plant shutdowns. *Quarterly Journal of Economics, 45*, 309–346.

Clark, C. (1940, [2nd ed. 1951]). *The conditions of economic progress*. Macmillan & Company, Ltd.

Clark, J. M. (1949). Common and disparate elements in national growth and decline. In *Problems in the study of economic growth*. Universities-National Bureau. http://www.nber.org/chapters/c9511

Cobb, C., & Douglas, P. H. (1928). A theory of production. *American Economic Review,* 18(Supplement), 139–165.

Coelli, T., Lauwers, L., & Van Huylenbroeck, G. (2007). Environmental efficiency measurement and the materials balance condition. *Journal of Productivity Analysis, 28*, 3–12.

Costanza, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., Paruelo, J., Raskin, R. G., Sutton, P., & Van Den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature, 387*, 253–260.

Costanza, R., Hart, M., Posner, S., & Talberth, J. (2009). *Beyond GDP: The need for new measures of progress* (Pardee Paper No. 4. Pardee Center for the Study of the Longer-Range Future). https://www.bu.edu/pardee/files/documents/PP-004-GDP.pdf

Criscuolo, C. (2015). Productivity is soaring at top firms and sluggish everywhere else. *Harvard Business Review*. https://hbr.org/2015/08/productivity-is-soaring-at-top-firms-and-sluggish-everywhere-else

Cruz Alvarez, J. L., & Rossi-Hansberg, E. (2021). *The economic geography of global warming* (NBER Working Paper 28466). https://www.nber.org/papers/w28466

Cummins, J. D., Weiss, M. A., Xie, X., & Zi, H. (2010). Economies of scope in financial services: A DEA efficiency analysis of the US insurance industry. *Journal of Banking & Finance, 34*, 1525–1539.

Dakpo, K. H., Jeanneaux, P., & Latruffe, L. (2016). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. *European Journal of Operational Research, 250*, 347–359.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *American Economic Review, 80*, 355–361.

Davis, H. S. (1947). *The industrial study of economic progress*. University of Pennsylvania Press.

Davis, H. S. (1955). *Productivity accounting*. University of Pennsylvania Press.

Davis, J. J. (1927). The problem of the worker displaced by machinery. *Monthly Labor Review, 25*, 32–34.

Davis, M. A., Ghent, A. C., & Gregory, J. M. (2021). *The work-at-home technology boon and its consequences* (NBER Working Paper 28461). https://www.nber.org/papers/w28461
Dayre, J. (1951). La Productivité Intégrale du Travail et sa Mesure. *Revue d'Économie Politique, 61*, 665–675.
Dean, J. (1941). *The relation of cost to output for a leather belt shop*. NBER https://www.nber.org/books/dean41-1
Dechezleprêtre, A., Nachtigall, D., & Stadler, B. (2020). *The effect of energy prices and environmental policy stringency on manufacturing employment in OECD countries: Sector- and firm-level evidence* (OECD Economics Department Working Papers No. 1625). https://doi.org/10.1787/899eb13f-en
Deloitte. (2020). *Feeling the heat?* https://www2.deloitte.com/content/dam/Deloitte/de/Documents/risk/DI_Feeling-the-heat-sustainability.pdf
Demmou, L., & Franco, G. (2020). *Do sound infrastructure governance and regulation affect productivity growth? New insights from firm-level data* (OECD Economics Department Working Papers No. 1609). https://doi.org/10.1787.410535403555
Denison, E. F. (1962). *The sources of economic growth in the United States and the alternatives before us*. New York: Committee for Economic Development, Supplementary Paper No. 13.
Denison, E. F. (1974). *Accounting for United States economic growth 1929–1969*. The Brookings Institution.
D'Erasmo, P. N., & Moscoso Boedo, H. J. (2012). Financial structure, informality and development. *Journal of Monetary Economics, 59*, 286–302.
de Roest, K., Ferrari, P., & Knickel, K. (2018). Specialization and economies of scale or diversification and economies of scope? Assessing different agricultural development pathways. *Journal of Rural Studies, 59*, 222–231.
Dias, D. A., Marques, C. R., & Richmond, C. (2016). Misallocation and productivity in the lead up to the Eurozone crisis. *Journal of Macroeconomics, 49*, 46–70.
Dieppe, A. (ed) (2020), *Global Productivity: Trends, Drivers, and Policies*. The World Bank. https://www.worldbank.org/en/research/publication/global-productivity
Diewert, W. E. (1993). The early history of price index research, Chapter 2 In W. E. Diewert & A. O. Nakamura (Eds.), *Essays in index number theory* (Vol 1). North-Holland.
Diewert, W. E. (in press a). The axiomatic or test approach to index number theory, Chapter 3. In *Consumer price index theory*. International Monetary Fund.
Diewert, W. E. (in press b). The economic approach to index number theory, Chapter 5. In *Consumer price index theory*. Washington: International Monetary Fund.

Drucker, P. F. (1954). *The practice of management*. Harper & Row.
Easterly, W., & Levine, R. (2001). It's not factor accumulation: Stylized facts and growth models. *World Bank Economic Review, 15*, 177–219.
Economist. (2016, June 25). *Special report: Artificial intelligence*. https://www.economist.com/special-report/2016/06/23/the-return-of-the-machinery-question
Economist. (2017, August 26). Artificial intelligence will create new kinds of work. https://www.economist.com/business/2017/08/26/artificial-intelligence-will-create-new-kinds-of-work
Economist. (2020a, March 26). *The coronavirus crisis thrusts corporate HR chiefs into the spotlight*. https://www.economist.com/business/2020/03/26/the-coronavirus-crisis-thrusts-corporate-hr-chiefs-into-the-spotlight
Economist. (2020b, September 17). *Special report: Business and climate change*. https://www.economist.com/special-report/2020/09/17/the-great-disrupter
Economist. (2020c, October 8). *The pandemic could give way to an era of rapid productivity growth*. https://www.economist.com/finance-and-economics/2020/12/08/the-pandemic-could-give-way-to-an-era-of-rapid-productivity-growth
Economist (2021, March 27). *Bright Side of the Moonshots*. https://www.economist.com/leaders/2021/03/27/bright-side-of-the-moonshots and "A New PhaRNAcopoeia" https://www.economist.com/briefing/2021/03/27/covid-19-vaccines-have-alerted-the-world-to-the-power-of-rna-therapies
Égert, B. (2016). Regulation, institutions and productivity: New macroeconomic evidence from OECD countries. *American Economic Review, 106*, 109–113.
Eldor, D., & Sudit, E. (1981). Productivity-based financial net income analysis. *Omega, 9*, 605–611.
Engström, G., Gars, J., Jaakkola, N., Lindahl, T., Spiro, D., & van Benthem, A. A. (2020). What policies address both the coronavirus crisis and the climate crisis? *Environmental and Resource Economics, 76*, 789–810.
Fabricant, S. (1940). *The output of manufacturing industries, 1899–1937*. National Bureau of Economic Research. http://www.nber.org/chapters/c6435
Fabricant, S. (1942). *Employment in manufacturing, 1899–1939*. National Bureau of Economic Research. https://www.nber.org/books-and-chapters/employment-manufacturing-1899-1939-analysis-its-relation-volume-production
Fagan, H. B. (1935). *American economic progress*. J. B. Lippincott Company.
Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A, General 120 (Part 3), 253–281.
Feigenbaum, J., & Gross, D. P. (2020). *Automation and the fate of young workers: Evidence from telephone operation in the early 20th century* (NBER Working Paper 28061). http://www.nber.org/papers/w28061

Fernald, J. G. (2015). Productivity and potential output before, during and after the great recession, Chapter 1. *NBER Macroeconomics Annual, 29*, 1–51.
Fernald, J., & Neiman, B. (2011). Growth accounting with misallocation: Or, doing less with more in Singapore. *American Economic Journal: Macroeconomics, 3*, 29–74.
Fernald, J. G., & Wang, J. C. (2016). Why has the cyclicality of productivity changed? What does it mean? *Annual Review of Economics, 8*, 465–496.
Field, A. J. (2003). The most technologically progressive decade of the century. *American Economic Review, 93*, 1399–1413.
Forster, P. M., Forster, H. I., Evans, M. J., Gidden, M. J., Jones, C. D., Keller, C. A., Lamboll, R.D., Quéré, C. L., Rogelj, J., Rosen, D., & Schleussner, C. F. (2020). Current and future global climate impacts resulting from COVID-19. *Nature Climate Change, 10*, 913–191.
Førsund, F. R. (2018). Productivity measurement and the environment, Chapter 8. In E. Grifell-Tatjé, C. A. K. Lovell & R. C. Sickles (Eds.), *The Oxford handbook of productivity analysis*. Oxford University Press.
Fourastié, J. (1951). *Machinisme et Bien-Être*. http://www.fourastie-sauvy.org
Fourastié, J. (1957). *Productivity, prices and wages* (European Productivity Agency Project No 235).
Freeman, D., Inklaar, R., & Diewert, W. E. (2021). Natural resources and missing inputs in international productivity comparisons. *Review of Income and Wealth, 67*, 1–17. https://doi.org/10.1111/roiw.12451
Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., & Le Quéré, C. (2020). Global carbon budget 2020. *Earth System Science Data, 12*, 3269–3340.
Fukuda, S., & Nakamura, J. (2011). Why did 'zombie' firms recover in Japan? *The World Economy, 34*, 1124–1137.
Gal, P., Nicoletti, G., Renault, T., Sorbe, S., & Timiliotis, C. (2019). *Digitalisation and productivity: In search of the holy grail* (OECD Economics Department Working Papers No. 1533). https://doi.org/10.1787/5080f4b6-en
Galbraith, J. K. (1955). *The great crash 1929*. Houghton Mifflin.
Galí, J., & van Rens, T. (2021). The vanishing procyclicality of labour productivity. *Economic Journal, 131*, 271–301.
Garcia Marin, A., Potlogea, A. V., Voigtländer, N., & Yang, Y. (2021). *Cities, productivity and trade* (NBER Working Paper 28309). https://www.nber.org/papers/w28309
García-Santana, M., Moral-Benito, E., Pijoan-Mas, J., & Ramos, R. (2020). Growing like Spain: 1995-2007. *International Economic Review, 61*, 383–416.
Gardiner, B. (2020). Coronavirus holds key lessons on how to fight climate change. *Yale Environment, 360*. https://e360.yale.edu/features/coronavirus-holds-key-lessons-on-how-to-fight-climate-change

Garnaut, R. (2019). *Super-power: Australia's low-carbon opportunity*. La Trobe University Press.

Georgescu-Roegen, N. (1951). The aggregate linear production function and its applications to von Neumann's economic model, Chapter IV. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation*. Wiley.

Gettins, M. (2020). The climate crisis will not wait. *Intereconomics, 55*, 364–365.

Gibbs, M., Mengel, F., & Siemroth, C. (2021). *Work from home & productivity: Evidence from personnel & analytics data on IT professionals* (Working Paper No. 2021-56). Becker-Friedman Institute https://bfi.uchicago.edu/working-paper/2021-56/

Gold, B. (1971). *Explorations in managerial economics: Productivity, costs, technology and growth*. Basic Books.

Gordon, R. J. (2016). *The rise and fall of American growth*. Princeton University Press.

Gosnell, G. K., List, J. A., & Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy, 128*, 1195–1233.

Goto, Y., & Wilbur, S. (2019). Unfinished business: Zombie firms among SME in Japan's lost decades. *Japan and the World Economy, 49*, 105–112.

Gouveia, A. F., & Osterhold, C. (2018). *Fear the walking dead: Zombie firms, spillovers and exit barriers* (OECD Productivity Working Paper No. 13). https://www.oecd-ilibrary.org/docserver/e6c6e51d-en.pdf?expires=1606775473&id=id&accname=guest&checksum=647C3FAD8C7F0DACF7749C641181C03D

Graetz, G., & Michaels, G. (2018). Robots at work. *Review of Economics and Statistics, 100*, 753–768.

Green Growth Knowledge Platform. (2016). *Measuring inclusive green growth at the country level* (GGKP Research Committee on Measurement & Indicators Working Paper 02/2016). https://www.greengrowthknowledge.org/research/measuring-inclusive-green-growth-country-level

Greene, K. (2000). Technological innovation and economic progress in the ancient world: M. I. Finley revisited. *Economic History Review, 53*, 29–59.

Grifell-Tatjé, E., & Lovell, C. A. K. (2003). The managers versus the consultants. *Scandinavian Journal of Economics, 105*, 119–138.

Grifell-Tatjé, E., & Lovell, C. A. K. (2015). *Productivity accounting: The economics of business performance*. Cambridge University Press.

Grifell-Tatjé, E., Lovell, C. A. K., & Sickles, R. C. (2018a). Overview of productivity analysis: History, issues and perspectives, Chapter 1. In E. Grifell-Tatjé, C. A. K. Lovell & R. C. Sickles (Eds.), *The Oxford handbook of productivity analysis*. Oxford University Press.

Grifell-Tatjé, E., Lovell, C. A. K., & Turon, P. (2018b). The business foundations of social economic progress. *Business Research Quarterly, 21*, 278–292.

Griliches, Z. (1957). Specification bias in estimates of production functions. *Journal of Farm Economics, 39*, 8–20.
Griliches, Z. (1996). The discovery of the residual: A historical note. *Journal of Economic Literature, 34*, 1324–1330.
Grosskopf, S., Hayes, K. J., Taylor, L. L., & Weber, W. L. (1999). Anticipating the consequences of school reform: A new use of DEA. *Management Science, 45*, 608–620.
Growitsch, C., & Wetzel, H. (2009). Testing for economies of scope in European railways: An efficiency analysis. *Journal of Transport Economics and Policy, 43*, 1–24.
Guerry, A. D., Polasky, S., Lubchenco, J., Chaplin-Kramer, R., Daily, G. C., Griffin, R., Ruckelshaus, M., Bateman, I. J., Duraiappah, A., Elmqvist, T., & Feldman, M. W. (2015). Natural capital and ecosystem services informing decisions: From promise to practice. *Proceedings of the National Academy of Sciences, 112*, 7348–7355.
Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics, 114*, 83–116.
Hamilton, E. J. (1944). Use and misuse of price history. *Journal of Economic History, 4*(Supplement), 47–60.
Heal, G., & Park, J. (2013). *Feeling the heat: Temperature, physiology & the wealth of nations* (NBER Working Paper 19725). http://www.nber.org/papers/w19725
Helm, D. (2020). The environmental impacts of the coronavirus. *Environmental and Resource Economics, 76*, 21–38.
Hepburn, C., O'Callaghan, B., Stern, N., Stiglitz, J., & Zenghelis, D. (2020). Will COVID-19 fiscal recovery packages accelerate or retard progress on climate change? *Oxford Review of Economic Policy, 36*, S359–S381.
Hermansen, M. (2020). *Anti-competitive and regulatory barriers in the United States Labour Market* (OECD Economics Department Working Papers No. 1627). https://doi.org/10.1787/38649656-en
Hoch, I. (1955). Estimation of production function parameters and testing for efficiency. *Econometrica, 23*, 325–326.
Hoch, I. (1962). Estimation of production function parameters combining time-series and cross-section data. *Econometrica, 30*, 34–53.
Hoffman, M., & Tadelis, S. (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy, 129*, 243–285.
Hsieh, C.-T. (2002). What explains the industrial revolution in East Asia? Evidence from the factor markets. *American Economic Review, 92*, 502–526.
Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics, 124*, 1403–1448.

Hsieh, C.-T., & Moretti, E. (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics, 11*, 1–39.
Hulten, C. R. (2001). Total factor productivity: A short biography, Chapter 1. In C. R. Hulten, E. R. Dean & M. J. Harper (Eds.), *New developments in productivity analysis*. University of Chicago Press.
Ichniowski, C., & Shaw, K. (2012). Insider econometrics, Chapter 7. In R. Gibbons & J. Roberts (Eds.), *The handbook of organizational economics*. Princeton University Press.
Ingham, H. (1961). Inter-firm comparison for management. *Productivity Measurement Review, 26*, 5–19.
Intergovernmental Panel on Climate Change (IPCC). (2014). *Climate Change 2014: Impacts, adaptation, and vulnerability*. https://www.ipcc.ch/report/ar5/wg2
Intergovernmental Panel on Climate Change (IPCC). (2018). *Global warming of 1.5 °C*. https://www.ipcc.ch/sr15/
International Labour Organization. (2020). *Consumer price index manual: Concepts and methods*. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_761444.pdf
International Monetary Fund. (2020). *World economic outlook: A long and difficult ascent*. https://www.imf.org/en/Publications/WEO/Issues/2020/09/30/world-economic-outlook-october-2020
Jerome, H. (1932). The measurement of productivity changes and the displacement of labor. *American Economic Review, 22*, 32–40.
Jerome, H. (1934). *Mechanization in industry*. National Bureau of Economic Research. https://www.nber.org/books/jero34-1
Johansen, L. (1972). *Production functions: An integration of micro and macro, short run and long run aspects*. North-Holland.
Johnson, H. T. (1975). Management accounting in an early integrated industrial: E. I. duPont de Nemours Powder Company, 1903–1912. *Business History Review, 49*, 184–204.
Johnson, H. T. (1978). Management accounting in an early multidivisional organization: General motors in the 1920s. *Business History Review, 52*, 490–517.
Jorgenson, D. W., & Griliches, Z. (1967). The explanation of productivity change. *Review of Economic Studies, 34*, 249–283.
Kantorovich, L. V. (1939 [1960]). Mathematical methods of organizing and planning production. Leningrad University. Translated in *Management Science*, *6*, 366–422.
Kaplan, G., Moll, B., & Violante, G. L. (2020). *The great lockdown and the big stimulus: Tracing the pandemic possibility frontier for the U.S.* (NBER Working Paper 27794). http://www.nber.org/papers/w27794

Kareiva, P. M., McNally, B. W., McCormick, S., Miller, T., & Ruckelshaus, M. (2015). Improving global environmental management with standard corporate reporting. *Proceedings of the National Academy of Sciences, 112*, 7375–7382.

Kendrick, J. W. (1956). *Productivity trends: Capital and labor* (NBER Occasional Paper 53). https://www.nber.org/books/kend56-1

Kendrick, J. W. (1961). *Productivity trends in the United States*. Princeton University Press.

Kendrick, J. W., & Grossman, E. S. (1980). *Productivity in the United States: Trends and cycles*. The Johns Hopkins University Press.

Keuschnigg, C., & Kogler, M. (2020). The Schumpeterian role of banks: Credit reallocation and capital structure. *European Economic Review, 121*, 103349.

Keynes, J. M. (1931). Economic possibilities for our grandchildren. In *Essays in persuasion*. W. W. Norton & Co. http://www.econ.yale.edu/smith/econ116a/keynes1.pdf

Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities, Chapter III. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation*. Cowles Commission for Research in Economics Monograph No. 13. Wiley.

Krugman, P. (1991a). *Geography and trade*. MIT Press.

Krugman, P. (1991b). Increasing returns and economic geography. *Journal of Political Economy, 99*, 483–499.

Krugman, P. (1998). What's new about the new economic geography? *Oxford Review of Economic Policy, 14*, 7–17.

Kuznets, S. (1934). *National Income, 1929–32, Senate Document No.* 124, 73rd US Congress, 2nd Session. https://fraser.stlouisfed.org/title/national-income-1929-1932-971

Lancet. (2020). Editorial: Climate and COVID-19: Converging crises, 397, 71.

Lazear, E. P., Shaw, K. L., & Stanton, C. T. (2015). The value of bosses. *Journal of Labor Economics, 33*, 823–861.

Lemann, N. (1991). *The promised land*. Vintage Books.

Lemos, R., Muralidharan, K., & Scur, D. (2021). *Personnel management and school productivity: Evidence from India* (NBER Working Paper 28336). https://www.nber.org/papers/w28336

Le Quéré, C., Jackson, R. B., Jones, M. W., Smith, A. J., Abernethy, S., Andrew, R. M., De-Gol, A. J., Willis, D. R., Shan, Y., Canadell, J. G., & Friedlingstein, P. (2020). Temporary reduction in daily global CO_2 emissions during the COVID-19 forced confinement. *Nature Climate Change, 10*, 647–653.

Li, P., & Ouyang, Y. (2020). Technical change and green productivity. *Environmental and Resource Economics, 76*, 271–298.

Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S. J., Feng, S., Zheng, B., Cui, D., Dou, X., Zhu, B., & Guo, R. (2020). Near-real-time monitoring of global CO_2 emissions reveals the effects of the COVID-19 pandemic. *Nature Communications, 11*, 5172.

Lomax, K. S. (1952). Cost curves for electricity generation. *Economica, 19*, 193–197.

Los Angeles Times. (1989). *Small farmers lose long legal battle over UC agriculture study*. https://www.latimes.com/archives/la-xpm-1989-09-07-mn-2481-story.html

Lovell, C. A. K. (2021). *The pandemic, the climate, and productivity* (Working Paper WP11/2021). School of Economics, University of Queensland https://ideas.repec.org/p/qld/uqcepa/165.html

Lubin, I. (1929a). Measuring the labor absorbing power of American industry. *Journal of the American Statistical Association, 24*, 27–32.

Lubin, I. (1929b). *The absorption of the unemployed by American Industry*. Brookings Institute Pamphlet Series I:3. Brookings Institution. https://babel.hathitrust.org/cgi/pt?id=uc1.$b96127&view=1up&seq=3.

Maddison, A. (2005). Measuring and interpreting world economic performance 1500–2001. *Review of Income and Wealth, 51*(1), 1–35.

Maddison, A. (2006). *The world economy*. OECD Development Centre Studies. https://www.oecd-ilibrary.org/development/the-world-economy_9789264104143-en

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadística, 4*, 209–242.

Marshall, A. (1887). The theory of business profits: Reply. *Quarterly Journal of Economics, 1*, 477–481.

Marshall, A. (1890 [9th Variorum ed. 1961]). *Principles of economics, Ninth (Variorum) Edition*. Macmillan and Co. Ltd.

Martin, P. L., & Olmstead, A. L. (1985). The agricultural mechanization controversy. *Science, 227*, 601–606.

Massell, B. F. (1967). Elimination of management bias from production functions fitted to cross-section data: A model and an application to African agriculture. *Econometrica, 35*, 495–508.

Maue, C. C., Burke, M., & Emerick, K. J. (2020). *Productivity dispersion and persistence among the world's most numerous firms* (NBER Working Paper 26924). http://www.nber.org/papers/w26924

McKinsey & Company. (2020a). *Addressing climate change in a post-pandemic world*. https://www.mckinsey.com/business-functions/sustainability/our-insights/addressing-climate-change-in-a-post-pandemic-world

McKinsey & Company. (2020b). *How a post-pandemic stimulus can both create jobs and help the climate*. https://www.mckinsey.com/business-functions/sustainability/our-insights/how-a-post-pandemic-stimulus-can-both-create-jobs-and-help-the-climate

McKinsey & Company. (2020c). *Confronting climate risk*. https://www.mckinsey.com/business-functions/sustainability/our-insights/confronting-climate-risk
McKinsey & Company. (2020d). COVID-19 and climate change expose dangers of unstable supply chains. https://mckinsey.com/business-functions/operations/our-insights/covid-19-and-climate-change-expose-dangers-of-unstable-supply-chains
McKinsey & Company. (2020e). *What's next for remote work: An analysis of 2,000 tasks, 800 jobs and nine countries*. https://www.mckinsey.com/featured-insights/future-of-work/whats-next-for-remote-work-an-analysis-of-2000-tasks-800-jobs-and-nine-countries
McKinsey & Company. (2020f). *How COVID-19 has pushed companies over the technology tipping point—and transformed business forever*. https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever
McKinsey & Company. (multiple dates). *Future of work*. https://www.mckinsey.com/featured-insights/future-of-work
Medical Journal of Australia. (2020). The 2020 special report of the MJA-lancet countdown on health and climate change: Lessons learnt from Australia's 'black summer', *213*, 490–492.
Mefford, R. N. (1986). Introducing management into the production function. *Review of Economics and Statistics, 68*, 96–104.
Mendershausen, H. (1938). On the significance of professor Douglas' production function. *Econometrica, 6*, 143–153.
Miller, D. M. (1984). Profitability = productivity + price recovery. *Harvard Business Review, 62*, 145–153.
Mills, F. C. (1932). *Economic tendencies in the United States: Aspects of pre-war and post-war changes*. http://papers.nber.org/books/mill32-1
Mokyr, J., Vickers, C., & Ziebarth, N. L. (2015). The history of technological anxiety and the future of economic growth: Is this time different? *Journal of Economic Perspectives, 29*, 31–50.
Morrison, C. J. (1992). Unraveling the productivity growth slowdown in the United States, Canada, and Japan: The effects of subequilibrium, scale economies and markups. *Review of Economics and Statistics, 74*, 381–393.
Moscoso Boedo, H. J., & Mukoyama, T. (2012). Evaluating the effects of entry regulations and firing costs on international income differences. *Journal of Economic Growth, 17*, 143–170.
Mundlak, Y. (1961). Empirical production function free of management bias. *Journal of Farm Economics, 43*, 44–56.
Myers, R. J. (1929). Occupational readjustment of displaced skilled workmen. *Journal of Political Economy, 37*, 473–489.

Nath, I. B. (2020). *The food problem and the aggregate productivity consequences of climate change* (NBER Working Paper 27297). http://www.nber.org/papers/w27297

New South Wales Treasury. (1999). *Profit composition analysis: A technique for linking productivity measurement & financial performance* (Office of Financial Management Research and Information Paper TRP 99-5). Sydney: New South Wales Treasury. www.treasury.nsw.gov.au/indexes/trpindex.html

Nguyen, B. H., Sickles, R. C., & Zelenyuk, V. (in press). Efficiency analysis with stochastic frontier models using Stata, Matlab and R. this *Volume*.

Norberg, J. (2020). *Open: The story of human progress*. Atlantic Books.

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences, 103*, 3510–3517.

Nordhaus, W. D., & Kokkelenberg, E. C. (1999). *Nature's numbers—Expanding the national accounts to include the environment*. National Academy Press. https://doi.org/10.17226/6374

Nordin, J. A. (1947). Note on a light plant's cost curves. *Econometrica, 15*, 231–235.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

O'Donnell, C. J. (2021). *Estimating the impacts of weather and climate change on agricultural productivity* (University of Queensland School of Economics CEPA Working Paper WP/03/2021). https://economics.uq.edu.au/files/24870/WP032021.pdf

OECD. (2001). *Measuring productivity OECD manual*. http://www.oecd.org/sdd/productivity-stats/2352458.pdf

OECD. (2014). *Making inclusive growth happen*. https://www.oecd.org/inclusive-growth/All-on-Board-Making-Inclusive-Growth-Happen.pdf

OECD. (2017). *Green growth indicators 2017*. https://doi.org/10.1787/9789264268586-en

OECD. (2018a). *For good measure: Advancing research on well-being metrics beyond GDP*. http://www.oecd.org/publications/for-good-measure-9789264307278-en.htm

OECD. (2018b). *China's progress towards green growth*. https://www.oecd.org/env/country-reviews/PR-China-Green-Growth-Progress-Report-2018b.pdf

OECD. (2019). Digitalisation and productivity: A story of complementarities, Chapter 2. In *OECD economic outlook Volume 2019 Issue 1*. OECD Publishing. https://doi.org/10.1787/5713bd7d-en

OECD. (2020a). *Building back better: A sustainable, resilient recovery after COVID-19*. http://www.oecd.org/coronavirus/policy-responses/building-back-better-a-sustainable-resilient-recovery-after-covid-19-52b869f5/

OECD. (2020b). *Making the green recovery work for jobs, income and growth*. https://read.oecd-ilibrary.org/view/?ref=136_136201-ctwt8p7qs5&title=Making-the-Green-Recovery-Work-for-Jobs-Income-and-Growth_
OECD. (2020c). *OECD economic outlook, preliminary version December* 2020c. https://doi.org/10.1787/39a88ab1-en
OECD. (2021). *Assessing the economic impacts of environmental policies*. https://www.oecd.org/environment/assessing-the-economic-impacts-of-environmental-policies-bf2fb156-en.htm
Ohanian, L. E. (2001). Why did productivity fall so much during the great depression? *American Economic Review, 91*, 34–38.
Osterhold, C., & Gouveia, A. F. (2020). Productivity, zombie firms and exit barriers in Portugal. *International Productivity Monitor, 38*, 29–49.
Panzar, J. C., & Willig, R. D. (1981). Economies of scope. *American Economic Review, 71*, 268–272.
Pearce, F. (2020). After the coronavirus, two sharply divergent paths on climate. *Yale Environment, 360* https://e360.yale.edu/features/after-the-coronavirus-two-sharply-divergent-paths-on-climate
Penrose, E. T. (1959). *The theory of the growth of the firm*. Basil Blackwell.
Peyrache, A., & Silva, M. C. A. (in press). Efficiency and productivity analysis from a system perspective: Historical overview. this *Volume*.
Pigou, A. C. (1920 [4th ed. 1960]). *The economics of welfare*. Macmillan and Company.
Porter, M. E. (1990). *The competitive advantage of nations*. Free Press.
Porter, M. E. (1991). America's Green strategy. *Scientific American, 264*, 168.
Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review, 76*, 77–90.
Porter, M. E. (2000). Location, competition and economic development: Local clusters in a global economy. *Economic Development Quarterly, 14*, 15–34.
Pulley, L. B., & Braunstein, Y. M. (1992). A Composite cost function for multiproduct firms with an application to economies of scope in banking. *Review of Economics and Statistics, 74*, 221–230.
Restuccia, D. (2019). Misallocation and aggregate productivity across time and space. *Canadian Journal of Economics/Revue Canadienne d'Économique, 52*, 5–32.
Restuccia, D., & Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics, 16*, 1–10.
Restuccia, D., & Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives, 31*, 151–174.
Rostovtzeff, M. I. (1926). *The social and economic history of the Roman Empire*. Oxford University Press.

Russell, R. R. (2018). Theoretical productivity indices, Chapter 4. In E. Grifell-Tatjé, C. A. K. Lovell & R. C. Sickles (Eds.), *The Oxford handbook of productivity analysis*. Oxford University Press.

Sachs, J. D., Mellinger, A. D., & Gallup, J. L. (2001). The geography of poverty and wealth. *Scientific American, 284*, 70–75.

Sandmo, A. (2015). The early history of environmental economics. *Review of Environmental Economics and Policy, 9*, 43–63.

Schivardi, F., Sette, E., & Tabellini, G. (2017). *Credit misallocation during the European financial crisis* (BIS Working Paper No. 669, Bank for International Settlements). https://www.bis.org/publ/work669.pdf

Schmookler, J. (1952). The changing efficiency of the American economy, 1869–1938. *Review of Economics and Statistics, 34*, 214–231.

Schreyer, P., & Diewert, W. E. (2014). Household production, leisure and living standards, Chapter 4. In D. W. Jorgenson, J. S. Landefeld & P. Schreyer (Eds.), *Measuring economic sustainability and progress*. University of Chicago Press.

Scientific American. (1982, September). *The mechanization of work*. 247

Scur, D., Sadun, R., Van Reenen, J., Lemos, R., & Bloom, N. (2021). *The world management survey at 18: Lessons and the way forward*. (NBER Working Paper 28524). https://www.nber.org/papers/w28524

Shephard, R. W. (1953). *Cost and production functions*. Princeton University Press.

Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton University Press.

Shephard, R. W. (1974). *Indirect production functions*. Verlag Anton Hain.

Sickles, R. C., Sun, K., & Triebs, T. P. (2021). The optimal use of management. *Economic Inquiry, 59*, 1346–1363.

Sickles, R. C., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency: Theory and practice*. Cambridge University Press.

Siegel, I. H. (1952). *Concepts and measurement of production and productivity*. Bureau of Labor Statistics, US Department of Labor.

Siegel, I. H. (1955). Aspects of productivity measurement and meaning, Chapter III. In European Productivity Agency (1955), *Productivity Measurement I. Concepts*. Paris: Organisation for European Economic Co-operation.

Slack. (2020). *Remote employee experience index*. https://slack.com/intl/en-au/blog/transformation/remote-employee-experience-index-launch

Smith, A. (1776 [1937]). *An inquiry into the nature and causes of the wealth of nations*. The Modern Library.

Social Indicators Research. (2011). National and personal well-being: Measuring the progress of societies, 102. https://link.springer.com/journal/11205/volumes-and-issues/102-1

Solow, R. M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics, 39*, 312–320.
Solow, R. M. (1987, July 12). We'd better watch out. *New York Times Book Review*.
Sorbe, S., Gal, P., Nicoletti, G., & Timiliotis, C. (2019). *Digital dividend: Policies to harness the productivity potential of digital technologies* (OECD Economic Policy Paper No. 26). https://www.oecd.org/economy/growth/digitalisation-productivity-and-inclusiveness/#d.en.520778
Spengler, J. J. (1949). Theories of socioeconomic growth. In *Problems in the study of economic growth*. Universities-National Bureau. https://www.nber.org/chapters/c9512
Spengler, J. J. (1959). Adam Smith's theory of economic growth parts I and II. *Southern Economic Journal, 25–26*, 397–415, 1–12.
Spratt, D., & Armistead, A. (2020). *COVID-19 climate lessons* (Discussion Paper, Breakthrough National Centre for Climate Restoration Australia). https://www.breakthroughonline.org.au/papers
Squires, B. M. (1917). Productivity and cost of labor in the lumber industry. *Monthly Labor Review, 5*, 66–79.
Staiger, D. (2020). What health care teaches us about measuring productivity in higher education, Chapter 2. In C. M. Hoxby & K. Strange (Eds.), *Productivity in higher education*. University of Chicago Press.
Stern, B. (1939). Labor productivity in the boot and shoe industry. *Monthly Labor Review, 48*, 271–292.
Stewart, E. (1922). Efficiency of American labor. *Monthly Labor Review, 15*, 1–12.
Stewart, E. (1924). Wastage of men. *Monthly Labor Review, 19*, 1–8.
Stigler, G. J. (1947). *Trends in output and employment*. National Bureau of Economic Research. http://www.nber.org/chapters/c7067
Stiglitz, J. E., Fitoussi, J.-P., & Durand, M. (2018). *Beyond GDP: Measuring what counts for economic and social performance*. OECD Publishing. https://doi.org/10.1787/9789264307292-en
Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009, December). *The measurement of economic performance and social progress revisited* (OFCE Centre de recherche en économie de sciences Po. Working Paper 2009-33). www.ofce.sciences-po.fr/pdf/dtravail/WP2009-33.pdf
Storz, M., Koetter, M., Setzer, R., & Westphal, A. (2017). *Do we want these two to tango? On zombie firms and stressed banks in Europe* (European Central Bank Working Paper No. 2104). https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2104.en.pdf
Survey of Current Business. (1972). *The measurement of productivity*, 52:5, Part II. Washington: US Department of Commerce, Bureau of Economic Analysis.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature, 49*, 326–365.

Tinbergen, J. (1942). Professor Douglas' production function. *Revue de l'Institut International de Statistique, 10*, 37–48.

Tintner, G. (1946). Applications of multivariate analysis to economic data. *Journal of the American Statistical Association, 41*, 472–500.

Tisdell, C. A. (2020). Economic, social and political issues raised by the COVID-19 pandemic. *Economic Analysis and Policy, 68*, 17–28.

Tol, R. S. J. (2009). The economic effects of climate change. *Journal of Economic Perspectives, 23*, 29–51.

Tridimas, G. (2019). The failure of ancient Greek growth: Institutions, culture and energy cost. *Journal of Institutional Economics, 15*, 327–350.

United Nations. (2009). *System of National Accounts 2008*. https://unstats.un.org/unsd/nationalaccount/sna2008.asp

United Nations. (2014). *System of environmental-economic accounting 2012*. https://unstats.un.org/unsd/envaccounting/seeaRev/SEEA_CF_Final_en.pdf

van Vuuren, D. P., van der Wijst, K.-I., Marsman, S., van den Berg, M., Hof, A. F., & Jones, C. D. (2020). The costs of achieving climate targets and the sources of uncertainty. *Nature Climate Change, 10*, 329–334.

Villegas, A., Molinos-Senante, M., Maziotis, A., & Sala-Garrido, R. (2020). Financial winners and losers since the privatization of the English and Welsh water and sewerage industry: A profit decomposition approach. *Urban Water Journal, 17*, 224–234.

Walker, F. A. (1887). The source of business profits. *Quarterly Journal of Economics, 1*, 265–288.

Wang, S. L., Ball, E., Nehring, R., Williams, R., & Chau, T. (2019). Impacts of climate change and extreme weather on U.S. agricultural productivity: Evidence and projection, Chapter 2. In W. Schlenker (Ed.), *Agricultural productivity and producer behavior*. University of Chicago Press.

White, K. D. (1956). The efficiency of Roman farming under the Empire. *Agricultural History, 30*, 85–89.

World Bank Group. (2020a). *Global economic prospects*. https://www.worldbank.org/en/publication/global-economic-prospects

World Bank Group. (2020b). *Doing business 2020b*. https://www.doingbusiness.org/en/doingbusiness

CHAPTER 2

Symmetric Decompositions of Aggregate Output and Labour Productivity Growth: On Levels, (Non-)Additivity, and Misallocation

Bert M. Balk

Introduction

The typical situation we will consider in this chapter is that of an economy consisting of a fixed number of industries. The mathematics, however, can also be applied to other situations, such as an industry consisting of a large number of firms (or establishments, or plants). In each case, we are looking at an ensemble of, more or less autonomous, consolidated production units. For an economy and its industries, their profit and loss accounts are provided by the National Accounts. For individual firms one must rely on business surveys carried out by official statistical agencies, or

B. M. Balk (✉)
Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands
e-mail: bert.m.balk@outlook.com

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_2

administrative data retrieved from company records, or some combination of sources.

The conventional measure of output is (nominal) value added (= revenue minus cost of intermediate inputs), since that can be compared and added over production units without running into double-counting problems. Aggregate value added, at the economy level, is called Gross Domestic Product (GDP). One is interested in the growth of real GDP and its industrial sources or components. Another important measure is labour productivity, real GDP per hour worked, as this can directly be related to various measures of welfare.

Several authors have developed decompositions of growth of GDP, total factor productivity, or labour productivity according to industrial components. Of older vintage are the decompositions derived in a theoretical framework based on neo-classical assumptions and/or continuous time. Classics are those by Hulten (1978), Gollop (1979), Jorgenson et al. (1987), Nordhaus (2002), and Stiroh (2002). Over the course of years, these decompositions have been reinvented, streamlined, or extended. More recently, we have seen a number of approaches outside the traditional, neo-classical framework. These include Tang and Wang (2004, 2015), Diewert (2010, 2015, 2016), Reinsdorf and Yuskavage (2010), and Dumagan (2013a, 2013b). Most of these decompositions are essentially asymmetric with respect to time and appear to contain terms that can be considered as mathematical artefacts, without economic meaning.

Balk (2014, 2021, Sections 6.3 and 6.4) provided symmetric decompositions for growth of GDP and labour productivity, based on the powerful instrument of the logarithmic mean. All these were formulated in terms of indices. In the present chapter, we return to this theory, simplifying the presentation by reformulating in terms of levels.[1]

Here is an outline of what is coming. Sections "Decomposition of Output Growth" and "Decomposition of (Simple) Labour Productivity Growth" discuss output growth and labour productivity growth, respectively. Section "Additivity and Misallocation" turns to the assumption of

[1] Decompositions of total factor productivity growth were discussed in terms of indices in Balk (2021, Chapter 7), and in terms of levels in Balk (2021, Chapter 8).

additivity and reviews against this backdrop several concepts of misallocation of labour as provided in the literature. Section "How to Overcome Non-additivity" discusses two recipes for enforcing additivity on non-additive data. A brief conclusion follows.

Decomposition of Output Growth

We consider an ensemble $\mathcal{K}$ of consolidated production units. The central place is occupied by the following relation, expressing the additivity of nominal value added,

$$\mathrm{VA}^{\mathcal{K}t} = \sum_{k \in \mathcal{K}} \mathrm{VA}^{kt}, \tag{2.1}$$

where t denotes an accounting period. It is assumed that $\mathrm{VA}^{kt} > 0$ for all the production units and all the time periods considered. For each production unit $k \in \mathcal{K}$ *real value added* is defined as

$$\mathrm{RVA}^{k}(t, b) \equiv \mathrm{VA}^{kt} / P^{k}_{\mathrm{VA}}(t, b); \tag{2.2}$$

that is, nominal value added at period t divided by (or, as one says, deflated by) a production unit-k-specific value-added based price index[2] for period t relative to a certain reference period b. Without loss of generality, it may be assumed that period b lies somewhere in the past and that production unit k already existed in period b. The functional form of the price indices may vary over the production units; in particular, the price indices may be direct or chained or of mixed form. The notation is chosen so as to emphasize that, unlike nominal value added, real value added is not observable, but the outcome of a function.

Rewriting the last expression yields

$$\mathrm{VA}^{kt} = P^{k}_{\mathrm{VA}}(t, b)\mathrm{RVA}^{k}(t, b) \ (k \in \mathcal{K}). \tag{2.3}$$

Nominal value added is here decomposed into a price component and a quantity component. It is assumed that $P^{k}_{\mathrm{VA}}(b, b) = 1$, so that $\mathrm{RVA}^{k}(b, b) = \mathrm{VA}^{kb}$; that is, in the reference period real value added is identical to nominal value added.

[2] On the construction of value-added based price and quantity indices, see Balk (2021, Chapter 2, Appendix B).

For the ensemble, which can be considered as an aggregate production unit, we have similarly

$$\mathrm{VA}^{\mathcal{K}t} = P_{\mathrm{VA}}^{\mathcal{K}}(t, b)\mathrm{RVA}^{\mathcal{K}}(t, b), \tag{2.4}$$

where $P_{\mathrm{VA}}^{\mathcal{K}}(t, b)$ is a value-added based price index for the ensemble $\mathcal{K}$ for period t relative to a certain reference period b. This index is supposed to be estimated from a sample of establishments and products. Its functional form may differ from those of the production unit-specific price indices. An assumption of technical nature is that all the price indices in expressions (2.3) and (2.4) are using the same reference period.

Substituting expressions (2.3) and (2.4) into (2.1) and dividing both sides by the price index for the aggregate, $P_{\mathrm{VA}}^{\mathcal{K}}(t, b)$, delivers a relation between real value added of the ensemble and real value added of the individual units,

$$\mathrm{RVA}^{\mathcal{K}}(t, b) = \sum_{k \in \mathcal{K}} \frac{P_{\mathrm{VA}}^{k}(t, b)}{P_{\mathrm{VA}}^{\mathcal{K}}(t, b)} \mathrm{RVA}^{k}(t, b). \tag{2.5}$$

Similarly, substituting expressions (2.3) and (2.4) into (2.1) but now dividing both sides by real value added of the ensemble, $\mathrm{RVA}^{\mathcal{K}}(t, b)$, delivers the dual relation between the price index for the ensemble and the individual price indices,

$$P_{\mathrm{VA}}^{\mathcal{K}}(t, b) = \sum_{k \in \mathcal{K}} \frac{\mathrm{RVA}^{k}(t, b)}{\mathrm{RVA}^{\mathcal{K}}(t, b)} P_{\mathrm{VA}}^{k}(t, b). \tag{2.6}$$

It is important to observe that, unlike nominal value added—see expression (2.1)—real value added is in general not additive. Moreover, as relative price changes $P_{\mathrm{VA}}^{k}(t, b)/P_{\mathrm{VA}}^{\mathcal{K}}(t, b)$ $(k \in \mathcal{K})$ do not necessarily add up to 1, real value added of the ensemble is not a weighted mean of real value added of the constituent production units either. Dually, the price index for the ensemble is not a weighted mean of the individual price indices. We will return to this issue in section "Decomposition of (Simple) Labour Productivity Growth".

Consider now another period t', say, prior to t. If the production units are industries, it is quite natural to assume that they exist in both periods and can be matched. If the production units are firms, we are considering the subset of firms available in both periods (i.e., the so-called continuing firms), which usually makes out the majority of firms in any ensemble.[3] Divide then both sides of expression (2.5) by aggregate real value added of period t', $\mathrm{RVA}^{\mathcal{K}}(t', b)$. After adding in numerator and denominator $\mathrm{RVA}^{k}(t', b)$ $(k \in \mathcal{K})$, we obtain

$$\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)} = \sum_{k\in\mathcal{K}} \left(\frac{P^{k}_{\mathrm{VA}}(t,b)}{P^{\mathcal{K}}_{\mathrm{VA}}(t,b)} \frac{\mathrm{RVA}^{k}(t',b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)} \right) \frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)}, \tag{2.7}$$

which expresses aggregate real-value-added change as a weighted sum of individual real-value-added changes. Notice that these weights generally do not add up to 1. The dual relation is

$$\frac{P^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{P^{\mathcal{K}}_{\mathrm{VA}}(t',b)} = \sum_{k\in\mathcal{K}} \left(\frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t,b)} \frac{P^{k}_{\mathrm{VA}}(t',b)}{P^{\mathcal{K}}_{\mathrm{VA}}(t',b)} \right) \frac{P^{k}_{\mathrm{VA}}(t,b)}{P^{k}_{\mathrm{VA}}(t',b)}. \tag{2.8}$$

Thus, aggregate value-added based price change is a weighted sum, but not a weighted mean, of individual price changes. In both expressions (2.7) and (2.8), the weights are a mixture of periods t' and t data.[4]

Evidently, the primal relations are more interesting than the dual ones. Using the definition of real value added once more, expression (2.7) can be rearranged as

$$\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)} = \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}} \left(\frac{P^{k}_{\mathrm{VA}}(t,b)/P^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{P^{k}_{\mathrm{VA}}(t',b)/P^{\mathcal{K}}_{\mathrm{VA}}(t',b)} \frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)} \right). \tag{2.9}$$

We see here that, apart from the relative price change term (i.e., the first term between brackets), real-value-added change of the ensemble is a nominal-value-added share-weighted mean of real-value-added changes of the individual production units. In terms of growth rates, this expression

[3] How to cope with a dynamic ensemble when measuring total factor productivity change is discussed in Balk (2021, Sections 7.7 and 7.8).

[4] Expressions (2.5) and (2.7) correspond to expression (2.8) of Dumagan (2014).

can be decomposed as

$$\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)} - 1 = \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}} \left(\frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)} - 1 \right) + \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}} \left(\frac{P^{k}_{\mathrm{VA}}(t,b)/P^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{P^{k}_{\mathrm{VA}}(t',b)/P^{\mathcal{K}}_{\mathrm{VA}}(t',b)} - 1 \right) \left(\frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)} - 1 \right) + \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}} \left(\frac{P^{k}_{\mathrm{VA}}(t,b)/P^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{P^{k}_{\mathrm{VA}}(t',b)/P^{\mathcal{K}}_{\mathrm{VA}}(t',b)} - 1 \right). \tag{2.10}$$

Corresponding with the Dumagan (2014) decomposition, the first term on the right-hand side of the equality sign could be called 'pure growth effect' (= weighted mean of individual growth rates), the second term 'growth-price interaction effect' (= weighted covariance of relative price changes and growth rates), and the third term 'relative price effect' (= weighted mean of relative price changes). The weights of the three terms are the same, namely nominal-value-added shares of period t'.

Combining the second and third terms delivers the simpler expression

$$\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)} - 1 = \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}} \left(\frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)} - 1 \right) + \sum_{k\in\mathcal{K}} \frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)} \left(\frac{P^{k}_{\mathrm{VA}}(t,b)}{P^{\mathcal{K}}_{\mathrm{VA}}(t,b)} - \frac{P^{k}_{\mathrm{VA}}(t',b)}{P^{\mathcal{K}}_{\mathrm{VA}}(t',b)} \right), \tag{2.11}$$

the last term of which corresponds to the 'price change effect' in the Dumagan (2016) decomposition. The weights of the two terms, however, are different. Moreover, the real-value-added based weights do not add up to 1.

Unfortunately, however, the decomposition in expression (2.9) is not unique. To see this, interchange in this expression the periods t and t'

and take reciprocals. This leads to an expression of the form

$$\left(\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)}\right)^{-1} = \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}} \left(\frac{P_{\mathrm{VA}}^{k}(t,b)/P_{\mathrm{VA}}^{\mathcal{K}}(t,b)}{P_{\mathrm{VA}}^{k}(t',b)/P_{\mathrm{VA}}^{\mathcal{K}}(t',b)} \frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)}\right)^{-1}, \tag{2.12}$$

which can be decomposed in terms of growth rates as

$$\begin{aligned}\left(\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)}\right)^{-1} - 1 = &\sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}} \left(\left(\frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)}\right)^{-1} - 1\right) \\ &+ \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}} \left(\left(\frac{P_{\mathrm{VA}}^{k}(t,b)/P_{\mathrm{VA}}^{\mathcal{K}}(t,b)}{P_{\mathrm{VA}}^{k}(t',b)/P_{\mathrm{VA}}^{\mathcal{K}}(t',b)}\right)^{-1} - 1\right) \\ &\left(\left(\frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)}\right)^{-1} - 1\right) \\ &+ \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}} \left(\left(\frac{P_{\mathrm{VA}}^{k}(t,b)/P_{\mathrm{VA}}^{\mathcal{K}}(t,b)}{P_{\mathrm{VA}}^{k}(t',b)/P_{\mathrm{VA}}^{\mathcal{K}}(t',b)}\right)^{-1} - 1\right).\end{aligned} \tag{2.13}$$

Whereas expression (2.9) is a weighted *arithmetic* mean of relative-price-change-corrected individual real-value-added ratios, expression (2.12) is an *harmonic* mean of the same entities. Moreover, the weights are different: the first uses period t' shares, but the second uses period t shares. Put another way, expressions (2.9) and (2.10) are forward-looking, whereas expressions (2.12) and (2.13) are backward-looking. The two decompositions of the same aggregate real-value-added change, $\mathrm{RVA}^{\mathcal{K}}(t,b)/\mathrm{RVA}^{\mathcal{K}}(t',b)$, are asymmetric with respect to time.

A *symmetric* decomposition can be obtained by taking, for instance, the geometric mean of the two asymmetric decompositions. Thus,

$$\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)} = \left(\frac{\sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}} \left(\frac{P_{\mathrm{VA}}^{k}(t,b)/P_{\mathrm{VA}}^{\mathcal{K}}(t,b)}{P_{\mathrm{VA}}^{k}(t',b)/P_{\mathrm{VA}}^{\mathcal{K}}(t',b)} \frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)}\right)}{\sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}} \left(\frac{P_{\mathrm{VA}}^{k}(t,b)/P_{\mathrm{VA}}^{\mathcal{K}}(t,b)}{P_{\mathrm{VA}}^{k}(t',b)/P_{\mathrm{VA}}^{\mathcal{K}}(t',b)} \frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)}\right)^{-1}}\right)^{1/2}. \tag{2.14}$$

This is, however, a very complex expression the parts of which are difficult to disentangle. A much simpler, symmetric decomposition of the aggregate real-value-added ratio can be obtained by employing the logarithmic mean. The aggregate nominal-value-added ratio can then be decomposed as

$$\ln\left(\frac{\mathrm{VA}^{\mathcal{K}t}}{\mathrm{VA}^{\mathcal{K}t'}}\right) = \sum_{k\in\mathcal{K}} \psi^k(t,t') \ln\left(\frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{kt'}}\right), \tag{2.15}$$

where

$$\psi^k(t,t') \equiv \frac{\mathrm{LM}\left(\frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}}, \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}}\right)}{\sum_{k\in\mathcal{K}} \mathrm{LM}\left(\frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}}, \frac{\mathrm{VA}^{kt'}}{\mathrm{VA}^{\mathcal{K}t'}}\right)} \; (k \in \mathcal{K})$$

and LM(.) is the logarithmic mean.[5] The coefficients $\psi^k(t,t')$ are (logarithmic) mean nominal-value-added shares, normalized so that they add up to 1. Notice that in the derivation of expression (2.15) no assumptions were involved.

Substituting expressions (2.3) and (2.4) into expression (2.15) yields

$$\ln\left(\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)}\right) + \ln\left(\frac{P_{\mathrm{VA}}^{\mathcal{K}}(t,b)}{P_{\mathrm{VA}}^{\mathcal{K}}(t',b)}\right) = \sum_{k\in\mathcal{K}} \psi^k(t,t') \ln\left(\frac{\mathrm{RVA}^k(t,b)}{\mathrm{RVA}^k(t',b)}\right) + \sum_{k\in\mathcal{K}} \psi^k(t,t') \ln\left(\frac{P_{\mathrm{VA}}^k(t,b)}{P_{\mathrm{VA}}^k(t',b)}\right). \tag{2.16}$$

[5] For any two strictly positive real numbers a and b, the logarithmic mean is defined by $\mathrm{LM}(a,b) \equiv (a-b)/\ln(a/b)$ if $a \neq b$ and $\mathrm{LM}(a,a) \equiv a$. It has the following properties: (1) $\min(a,b) \leq \mathrm{LM}(a,b) \leq \max(a,b)$; (2) $\mathrm{LM}(a,b)$ is continuous; (3) $\mathrm{LM}(\lambda a, \lambda b) = \lambda \mathrm{LM}(a,b)$ $(\lambda > 0)$; (4) $\mathrm{LM}(a,b) = \mathrm{LM}(b,a)$; (5) $(ab)^{1/2} \leq \mathrm{LM}(a,b) \leq (a+b)/2$; (6) $\mathrm{LM}(a,1)$ is concave. See Balk (2008, 134–136) for details.

Combining the price terms, and using the fact that the coefficients $\psi^k(t,t')$ add up to 1, then delivers

$$\ln\left(\frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t',b)}\right) = \sum_{k\in\mathcal{K}} \psi^k(t,t') \ln\left(\frac{\mathrm{RVA}^{k}(t,b)}{\mathrm{RVA}^{k}(t',b)}\right) + \sum_{k\in\mathcal{K}} \psi^k(t,t') \ln\left(\frac{P^k_{\mathrm{VA}}(t,b)/P^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{P^k_{\mathrm{VA}}(t',b)/P^{\mathcal{K}}_{\mathrm{VA}}(t',b)}\right). \tag{2.17}$$

Here, the aggregate real-value-added ratio appears to be the product of two components, a weighted geometric mean of individual real-value-added ratios and a weighted geometric mean of relative price changes. Expression (2.17) may directly be compared to expressions (2.10) and (2.13) by observing that, if $a \approx 1$ then $\ln a \approx a - 1$. Thus, the logarithms of ratios may be interpreted as growth rates. The most striking feature of expression (2.17) is then that *it does not contain an interaction term*. Such a term is an artefact, materializing only in asymmetric decompositions.

The relative-price-change term vanishes if and only if

$$\ln\left(\frac{P^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{P^{\mathcal{K}}_{\mathrm{VA}}(t',b)}\right) = \sum_{k\in\mathcal{K}} \psi^k(t,t') \ln\left(\frac{P^k_{\mathrm{VA}}(t,b)}{P^k_{\mathrm{VA}}(t',b)}\right). \tag{2.18}$$

This equality means that the aggregate price index for period t relative to period t', $P^{\mathcal{K}}_{\mathrm{VA}}(t,b)/P^{\mathcal{K}}_{\mathrm{VA}}(t',b)$, is a Sato-Vartia index of the individual price indices $P^k_{\mathrm{VA}}(t,b)/P^k_{\mathrm{VA}}(t',b)$ $(k \in \mathcal{K})$ (see Dumagan & Balk, 2016). A trivial case materializes when the same deflator is used for all the units and the ensemble; that is, when $P^k_{\mathrm{VA}}(t,b) = P^{\mathcal{K}}_{\mathrm{VA}}(t,b)$ $(k \in \mathcal{K})$.

In general, however, the relative-price-change term will not vanish. Even if all the individual production units face the same input and output prices, compositional differences between the units are responsible for differences in aggregate price developments.

Decomposition of (Simple) Labour Productivity Growth

We now turn to the decomposition of (simple) labour productivity change. Let L^{kt} denote the total quantity of labour input, measured in some common unit (say, hours worked), of production unit $k \in \mathcal{K}$ in period t. It is fairly natural to assume that

$$L^{\mathcal{K}t} = \sum_{k\in\mathcal{K}} L^{kt}, \tag{2.19}$$

where $L^{\mathcal{K}t}$ denotes the total labour input quantity of the ensemble. Put otherwise, we are assuming that an hour worked in production unit k is the same as an hour worked in production unit k' $(k, k' \in \mathcal{K})$.

Simple (value-added based) labour productivity is defined as real value added divided by labour input quantity (Balk 2021, 120); thus, for the individual production units as

$$\mathrm{SLPROD}^{k}_{\mathrm{VA}}(t, b) \equiv \mathrm{RVA}^{k}(t, b)/L^{kt} \;\; (k \in \mathcal{K}), \tag{2.20}$$

and for the ensemble as

$$\mathrm{SLPROD}^{\mathcal{K}}_{\mathrm{VA}}(t, b) \equiv \mathrm{RVA}^{\mathcal{K}}(t, b)/L^{\mathcal{K}t}. \tag{2.21}$$

The growth rate of aggregate simple labour productivity, going from period t' to period t, is then obtained by considering

$$\ln\left(\frac{\mathrm{SLPROD}^{\mathcal{K}}_{\mathrm{VA}}(t, b)}{\mathrm{SLPROD}^{\mathcal{K}}_{\mathrm{VA}}(t', b)}\right) = \ln\left(\frac{\mathrm{RVA}^{\mathcal{K}}(t, b)}{\mathrm{RVA}^{\mathcal{K}}(t', b)}\right) - \ln\left(\frac{L^{\mathcal{K}t}}{L^{\mathcal{K}t'}}\right). \tag{2.22}$$

The first term on the right-hand side of this expression can be decomposed according to expression (2.17). The second term can be decomposed, like expression (2.15), as

$$\ln\left(\frac{L^{\mathcal{K}t}}{L^{\mathcal{K}t'}}\right) = \sum_{k\in\mathcal{K}} \lambda^{k}(t, t') \ln\left(\frac{L^{kt}}{L^{kt'}}\right), \tag{2.23}$$

where

$$\lambda^{k}(t, t') \equiv \frac{\mathrm{LM}\left(\frac{L^{kt}}{L^{\mathcal{K}t}}, \frac{L^{kt'}}{L^{\mathcal{K}t'}}\right)}{\sum_{k\in\mathcal{K}} \mathrm{LM}\left(\frac{L^{kt}}{L^{\mathcal{K}t}}, \frac{L^{kt'}}{L^{\mathcal{K}t'}}\right)} \;\; (k \in \mathcal{K})$$

are (logarithmic) mean labour shares, normalized so that they add up to 1.

Using the definition of simple labour productivity as given by expression (2.20), and employing the fact that $\sum_{k\in\mathcal{K}}(\psi^k(t,t') - \lambda^k(t,t')) = 0$, it turns out that expression (2.22) may be written as

$$\ln\left(\frac{\text{SLPROD}^{\mathcal{K}}_{\text{VA}}(t,b)}{\text{SLPROD}^{\mathcal{K}}_{\text{VA}}(t',b)}\right) = \sum_{k\in\mathcal{K}}\psi^k(t,t')\ln\left(\frac{\text{SLPROD}^k(t,b)}{\text{SLPROD}^k(t',b)}\right) + \sum_{k\in\mathcal{K}}\psi^k(t,t')\ln\left(\frac{P^k_{\text{VA}}(t,b)/P^{\mathcal{K}}_{\text{VA}}(t,b)}{P^k_{\text{VA}}(t',b)/P^{\mathcal{K}}_{\text{VA}}(t',b)}\right) + \sum_{k\in\mathcal{K}}\psi^k(t,t')\left(1 - \frac{\lambda^k(t,t')}{\psi^k(t,t')}\right)\ln\left(\frac{L^{kt}/L^{\mathcal{K}t}}{L^{kt'}/L^{\mathcal{K}t'}}\right). \tag{2.24}$$

Hence, the growth rate of aggregate simple labour productivity consists of three main parts: a weighted mean of individual labour productivity growth rates, a weighted mean of relative price changes, and a covariance of labour share growth rate and the excess of labour share over value-added share, respectively. In all these parts, the weights are the same: (normalized logarithmic) mean nominal-value-added shares over the two periods considered. Notice that the covariance term vanishes if the labour shares of the production units coincide with their value-added shares. Notice further that expression (2.24) is not an approximation, but an identity, and that there were no assumptions involved in the derivation.

If all the deflators are transitive, so that the dependence on reference period b vanishes, then expression (2.24) corresponds to expression [4] of Dumagan and Balk (2016). An asymmetric, base-period (t') weighted variant was obtained by Diewert (2016, expression [1.9]). Diewert observed that, empirically, the relative-price-changes factor appeared to be insignificant.[6] But this does not come as a surprise. Recall our expression (2.18) and notice that, *to the first order*, the relation expressed there always holds since the inputs and outputs of the ensemble are the union of the inputs and outputs of all the individual production units.

The third term on the right-hand side of expression (2.24) deserves closer attention. Even if the individual labour productivities do not change, and there is no relative price change, then change of labour

[6] Though individual components appeared to be quite large for some industries in particular years.

shares causes aggregate labour productivity change. Now, labour shares add up to 1, and thus, a change of the labour share of a certain production unit k goes with a change of the labour share of at least one other unit k'. This can be made explicit by noticing that expression (2.23) implies that

$$\ln\left(\frac{L^{kt}/L^{\mathcal{K}t}}{L^{kt'}/L^{\mathcal{K}t'}}\right) = - \sum_{k' \in \mathcal{K}, k' \neq k} \frac{\lambda^{k'}(t,t')}{\lambda^{k}(t,t')} \ln\left(\frac{L^{k't}/L^{\mathcal{K}t}}{L^{k't'}/L^{\mathcal{K}t'}}\right) \ (k \in \mathcal{K}). \quad (2.25)$$

Hence, expression (2.24) can alternatively be written as

$$\begin{aligned} \ln\left(\frac{\mathrm{SLPROD}^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{\mathrm{SLPROD}^{\mathcal{K}}_{\mathrm{VA}}(t',b)}\right) &= \sum_{k \in \mathcal{K}} \psi^{k}(t,t') \ln\left(\frac{\mathrm{SLPROD}^{k}(t,b)}{\mathrm{SLPROD}^{k}(t',b)}\right) \\ &+ \sum_{k \in \mathcal{K}} \psi^{k}(t,t') \ln\left(\frac{P^{k}_{\mathrm{VA}}(t,b)/P^{\mathcal{K}}_{\mathrm{VA}}(t,b)}{P^{k}_{\mathrm{VA}}(t',b)/P^{\mathcal{K}}_{\mathrm{VA}}(t',b)}\right) \\ &- \sum_{k \in \mathcal{K}} \sum_{k' \in \mathcal{K}, k' \neq k} \psi^{k}(t,t') \left(1 - \frac{\lambda^{k}(t,t')}{\psi^{k}(t,t')}\right) \frac{\lambda^{k'}(t,t')}{\lambda^{k}(t,t')} \\ &\ln\left(\frac{L^{k't}/L^{\mathcal{K}t}}{L^{k't'}/L^{\mathcal{K}t'}}\right). \end{aligned} \quad (2.26)$$

In this way, the roles played by the labour shares of all the production units are made explicit. I believe this corresponds to the intuition underlying the decomposition method proposed by Baldwin and Willox (2016). The third term on the right-hand side of expression (2.24) considers (labour) reallocation from the perspective of each individual production unit k. Likewise, the third term on the right-hand side of expression (2.26) considers this from the perspective of all the other firms $k' \neq k$. There is no need to make a choice here.

Additivity and Misallocation

As noted in section "Decomposition of Output Growth", real value added is in general not additive. Additivity of real value added holds if and only if

$$\mathrm{RVA}^{\mathcal{K}}(t,b) = \sum_{k\in\mathcal{K}} \mathrm{RVA}^{k}(t,b). \tag{2.27}$$

Basically, this means that the real value added produced by production unit k is made of the same 'stuff' as the real value added produced by unit k' ($k, k' \in \mathcal{K}$), and thus, these 'quantities' may be added together. Put otherwise, all the product differences between the production units are neglected.

Given the definitions of aggregate and production unit-specific real value added, expression (2.27) may be replaced by

$$\left(P_{\mathrm{VA}}^{\mathcal{K}}(t,b)\right)^{-1} = \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{kt}}{\mathrm{VA}^{\mathcal{K}t}} \left(P_{\mathrm{VA}}^{k}(t,b)\right)^{-1}; \tag{2.28}$$

that is, the price index for the ensemble is a current-period-nominal-value-added weighted harmonic mean of the price indices for the individual production units (*aka* a Paasche index). This is of course a very severe restriction.

The virtue of assuming additivity is that aggregate simple labour productivity then takes on a simple form. Based on expression (2.27), we obtain

$$\mathrm{SLPROD}_{\mathrm{VA}}^{\mathcal{K}}(t,b) = \sum_{k\in\mathcal{K}} \mathrm{SLPROD}_{\mathrm{VA}}^{k}(t,b) \frac{L^{kt}}{L^{\mathcal{K}t}}; \tag{2.29}$$

that is, a aggregate simple labour productivity is a labour share-weighted arithmetic mean of individual simple labour productivities (recall the additivity of labour in expression (2.19)). Expression (2.27) can also be reformulated as

$$\mathrm{RVA}^{\mathcal{K}}(t,b) = \sum_{k\in\mathcal{K}} \mathrm{SLPROD}_{\mathrm{VA}}^{k}(t,b) L^{kt}. \tag{2.30}$$

This provides a convenient starting point for a discussion of the concept of *misallocation*. The basic idea behind this concept, of which no unequivocal definition is available in the literature, seems to be that an expression such as (2.30) is considered as representation of a production function with neo-classical traits.[7]

Expression (2.30) specifically pictures the production process of aggregate real value added as a linear one in which each firm, with given size L^{kt}, has selected an element of the set of productivities $\{\mathrm{SLPROD}_{\mathrm{VA}}^{k'}(t,b) \mid k' \in \mathcal{K}\}$. Evidently, the actual allocation of firms and productivities (or 'productivity shocks', as they are called in the literature) is not necessarily optimal. On the assumption that productivities can indeed be selected frictionless, maximal real value added would be

$$\sum_{k\in\mathcal{K}}\left(\max_{k'\in\mathcal{K}} \mathrm{SLPROD}_{\mathrm{VA}}^{k'}(t,b)\right)L^{kt} = \left(\max_{k'\in\mathcal{K}} \mathrm{SLPROD}_{\mathrm{VA}}^{k'}(t,b)\right)L^{\mathcal{K}t}. \quad (2.31)$$

Call this $\mathrm{RVAMAX}^{\mathcal{K}}(t,b)$. The alternative interpretation is that this is the aggregate real value added that could be obtained if the total labour supply shifts frictionless to the production unit exhibiting the highest productivity.[8]

The *(labour) allocation discrepancy* may then be defined as the ratio of actual to maximal real value added; that is,

$$\mathrm{LAD}_{\mathrm{VA}}^{\mathcal{K}}(t,b) \equiv \frac{\mathrm{RVA}^{\mathcal{K}}(t,b)}{\mathrm{RVAMAX}^{\mathcal{K}}(t,b)}, \quad (2.32)$$

the maximum value of which is 1. Substituting expressions (2.30) and (2.31), the labour allocation discrepancy can be expressed as

$$\mathrm{LAD}_{\mathrm{VA}}^{\mathcal{K}}(t,b) = \sum_{k\in\mathcal{K}} \frac{\mathrm{SLPROD}_{\mathrm{VA}}^{k}(t,b)}{\max_{k'\in\mathcal{K}} \mathrm{SLPROD}_{\mathrm{VA}}^{k'}(t,b)} \frac{L^{kt}}{L^{\mathcal{K}t}}; \quad (2.33)$$

that is, a labour share-weighted arithmetic mean of relative simple labour productivities. The relative gain from a better, even optimal, allocation of

[7] For instance, Hopenhayn (2014) considered, reformulated in our notation, $\sum_{k\in\mathcal{K}} \mathrm{SLPROD}_{\mathrm{VA}}^{k}(t,b)(L^{kt})^{\rho}$ with $\rho < 1$. However, such a model is difficult to reconcile with an accounting identity like expression (2.30).

[8] The Hopenhayn (2014) model delivered $\max_{L^{kt},k\in\mathcal{K}}\{\sum_{k\in\mathcal{K}} \mathrm{SLPROD}_{\mathrm{VA}}^{k}(t,b)(L^{kt})^{\rho} \mid \sum_{k\in\mathcal{K}} L^{kt} = L^{\mathcal{K}t}\} = \left(\sum_{k'\in\mathcal{K}}(\mathrm{SLPROD}_{\mathrm{VA}}^{k'}(t,b))^{1/(1-\rho)}\right)^{(1-\rho)}(L^{\mathcal{K}t})^{\rho}$.

labour input is inversely related to the allocation discrepancy, as

$$\frac{\mathrm{RVAMAX}^{\mathcal{K}}(t,b)}{\mathrm{RVA}^{\mathcal{K}}(t,b)} - 1 = \frac{1}{\mathrm{LAD}^{\mathcal{K}}_{\mathrm{VA}}(t,b)} - 1. \tag{2.34}$$

In a recent study, Gu (2019) proposed the difference between weighted and unweighted means of the individual productivities,

$$\sum_{k\in\mathcal{K}} \frac{L^{kt}}{L^{\mathcal{K}t}} \mathrm{SLPROD}^{k}_{\mathrm{VA}}(t,b) - \sum_{k\in\mathcal{K}} \frac{1}{\#(\mathcal{K})} \mathrm{SLPROD}^{k}_{\mathrm{VA}}(t,b), \tag{2.35}$$

where $\#(\mathcal{K})$ denotes the number of firms in the ensemble, as a measure of allocative efficiency. This increases if more productive firms increase their share of labour resources. The maximum value is

$$\max_{k\in\mathcal{K}} \mathrm{SLPROD}^{k}_{\mathrm{VA}}(t,b) - \sum_{k\in\mathcal{K}} \frac{1}{\#(\mathcal{K})} \mathrm{SLPROD}^{k}_{\mathrm{VA}}(t,b), \tag{2.36}$$

which suggests

$$\frac{\mathrm{RVA}^{\mathcal{K}}(t,b) - \left(\sum_{k\in\mathcal{K}} \frac{1}{\#(\mathcal{K})} \mathrm{SLPROD}^{k}_{\mathrm{VA}}(t,b)\right) L^{\mathcal{K}t}}{\mathrm{RVAMAX}^{\mathcal{K}}(t,b) - \left(\sum_{k\in\mathcal{K}} \frac{1}{\#(\mathcal{K})} \mathrm{SLPROD}^{k}_{\mathrm{VA}}(t,b)\right) L^{\mathcal{K}t}} \tag{2.37}$$

as an alternative measure of labour allocation discrepancy.

Microdata researchers, working with ensembles consisting of large numbers of individual firms or plants, are usually intrigued by the large dispersion of (labour) productivities and the large dispersion of firm sizes.[9] Given that, under the assumption of additivity, the 'stuff' produced by unit k is exchangeable to the 'stuff' produced by unit k', why is there not just one big production unit and, hence, one single productivity figure?

Thus, which causes are responsible for the empirical productivity dispersion? And why has labour supply, that is, the total labour input available to a particular ensemble, not migrated to the production unit with the largest productivity? These are some of the questions being considered

[9] For an overview of the issues, including a research agenda, see Bartelsman and Wolf (2018).

in OECD's MultiProd project (Berlingieri et al., 2017). The discussion in this chapter will be restricted to a number of measurement issues.

An obvious approach is to split the ensemble $\mathcal{K}$ into a number of disjunct sub-ensembles, $\mathcal{K}_d$ $(d = 1, ..., D)$. Then, the right-hand side of expression (2.29) can be written as

$$\text{SLPROD}_{\text{VA}}^{\mathcal{K}}(t, b) = \sum_{d=1}^{D} \frac{L^{\mathcal{K}_d t}}{L^{\mathcal{K} t}} \text{SLPROD}_{\text{VA}}^{\mathcal{K}_d}(t, b). \tag{2.38}$$

Thus, simple labour productivity of the entire ensemble is a weighted mean of the productivities of the sub-ensembles, the weights being sub-ensemble labour shares (notice that, by disjunctivity of the sub-ensembles, $L^{\mathcal{K}t} = \sum_{d=1}^{D} L^{\mathcal{K}_d t}$). The extent to which the productivity of the ensemble is dominated by sub-ensemble d is then captured by the ratio

$$\frac{L^{\mathcal{K}_d t}}{L^{\mathcal{K} t}} \frac{\text{SLPROD}_{\text{VA}}^{\mathcal{K}_d}(t, b)}{\text{SLPROD}_{\text{VA}}^{\mathcal{K}}(t, b)} \quad (d = 1, ..., D), \tag{2.39}$$

which can be presented as a percentage. The sub-ensembles could be size deciles (where, for instance, size is measured by sales). Then, the ratio in expression (2.39) for $d = D$ provides the extent to which the top decile dominates the productivity of the ensemble. As some theory predicts that resources flow to the largest production units, the larger this ratio the less misallocation there is.

The productivity dispersion in the ensemble itself may be measured by the (square root of the) (weighted) variance of the individual productivities,[10]

$$\begin{aligned} \text{var}\,(\text{SLPROD}_{\text{VA}}^{\mathcal{K}}(t, b)) \equiv \sum_{k \in \mathcal{K}} \frac{L^{kt}}{L^{\mathcal{K}t}} \Big(& \text{SLPROD}_{\text{VA}}^{k}(t, b) \\ & - \text{SLPROD}_{\text{VA}}^{\mathcal{K}}(t, b) \Big)^2. \end{aligned} \tag{2.40}$$

[10] An alternative is the interquartile range, which seems to be less sensitive to outliers. See Foster et al. (2021).

In the same way, the variances for each of the sub-ensembles $\mathcal{K}_d$ ($d = 1, \ldots, D$) may be calculated. Their relation is given by

$$\begin{aligned}\text{var}\,(\text{SLPROD}_{\text{VA}}^{\mathcal{K}}(t,b)) = &\sum_{d=1}^{D} \frac{L^{\mathcal{K}_d t}}{L^{\mathcal{K}t}} \text{var}\,(\text{SLPROD}_{\text{VA}}^{\mathcal{K}_d}(t,b)) \\ &+ \sum_{d=1}^{D} \frac{L^{\mathcal{K}_d t}}{L^{\mathcal{K}t}} \Big(\text{SLPROD}_{\text{VA}}^{\mathcal{K}_d}(t,b) \\ &- \text{SLPROD}_{\text{VA}}^{\mathcal{K}}(t,b)\Big)^2.\end{aligned} \tag{2.41}$$

Whereas expression (2.38) decomposes the first moment of the productivity distribution, expression (2.41) does the same with the second moment. The right-hand side of the latter expression, however, consists of *two* main components. The first is a labour share-weighted mean of sub-ensemble variances and the second is the variance of the sub-ensemble productivities. Put technically, the components concern within and between variance, respectively.

Like expression (2.39), the extent to which the productivity variance of the ensemble is dominated by sub-ensemble d is captured by the ratio

$$\frac{L^{\mathcal{K}_d t}}{L^{\mathcal{K}t}} \frac{\text{var}\,(\text{SLPROD}_{\text{VA}}^{\mathcal{K}_d}(t,b))}{\text{var}\,(\text{SLPROD}_{\text{VA}}^{\mathcal{K}}(t,b))} \quad (d = 1, \ldots, D), \tag{2.42}$$

which can also be presented as a percentage. Defining the sub-ensembles as size quantiles, Berlingieri et al. (2017) proposed these ratios for $d = 1$ and $d = D$ as signalling misallocation. However, the relationship with theory appears to be weak.

Finally, almost any study of productivity dispersion is hampered by the fact that at the level of individual production units specific prices are unavailable, so that theoretically required unit-specific deflators must be replaced by available ensemble-specific deflators—which implies additivity. Put otherwise, instead of (now called) *physical* productivities $\text{SLPROD}_{\text{VA}}^{k}(t,b)$ ($k \in \mathcal{K}$) one calculates *revenue* productivities,

$$\text{SRLPROD}_{\text{VA}}^{k}(t,b) \equiv \frac{\text{VA}^{kt}/P_{\text{VA}}^{\mathcal{K}}(t,b)}{L^{kt}}$$

$$= \text{SLPROD}^{k}_{\text{VA}}(t,b)\frac{P^{k}_{\text{VA}}(t,b)}{P^{\mathcal{K}}_{\text{VA}}(t,b)} \ (k \in \mathcal{K}). \qquad (2.43)$$

As the second line shows, revenue productivity equals physical productivity times the relative value-added price of the production unit. Some theory predicts that in a market where resources are efficiently allocated the dispersion of revenue productivity (as an empirical approximation of marginal revenue product) would be zero. Prominent in this line of thought is the much quoted article of Hsieh and Klenow (2009). However, Haltiwanger et al. (2018) have demonstrated that this theory requires a fair number of unrealistically strong assumptions. Equal revenue productivity among producers in a certain market appears to be not necessarily a sign of efficient allocation of resources. And reversely, an allocation that appears to be efficient leads not necessarily to equal revenue productivities. Moreover, from an empirical point of view the relation between revenue productivity dispersion (variance) and physical productivity dispersion (variance) is not very simple.[11] All in all, if and how both measures can be related to the misallocation issue is a topic of research. The reader is referred to De Loecker and Syverson (2021, Section 6.3), for a description of the state of the art.

How to Overcome Non-additivity

The non-additivity of real value added is generally considered a nuisance for users of National Accounts since it gives rise to non-allocatable residuals $\text{RVA}^{\mathcal{K}}(t,b) - \sum_{k\in\mathcal{K}} \text{RVA}^{k}(t,b)$. Several recipes have been offered to overcome this. I discuss two recent ones.

The conventional approach consists in deflating by annually chained Paasche price indices; that is

$$P^{\mathcal{K}}_{\text{VA}}(t,b) \equiv \prod_{\tau=b+1}^{t} P^{P\mathcal{K}}_{\text{VA}}(\tau,\tau-1) \qquad (2.44)$$

[11] As noted, revenue productivity equals physical productivity times a relative price. Consider two stochastic variables X and Y. It appears that $\text{var}(XY) = \text{cov}(X^2, Y^2) + \text{E}X^2\,\text{E}Y^2 - (\text{E}X)^2(\text{E}Y)^2[1 + \text{cov}(X,Y)/\text{E}X\,\text{E}Y]^2$, where E denotes mean, var denotes variance, and cov denotes covariance. Thus, there is no simple relation between $\text{var}(XY)$ and $\text{var}(X)$ or $\text{var}(Y)$.

$$P_{\text{VA}}^{k}(t,b) \equiv \prod_{\tau=b+1}^{t} P_{\text{VA}}^{Pk}(\tau,\tau-1)\ (k\in\mathcal{K}), \tag{2.45}$$

where the superscript P denotes Paasche. Using expressions (2.3) and (2.4) and the fact that Laspeyres-Paasche index pairs satisfy the Product Test, we see that the conventional approach can equivalently be described by the following system of real values

$$\text{RVA}^{\mathcal{K}}(t,b) \equiv \text{VA}^{\mathcal{K}b} \prod_{\tau=b+1}^{t} Q_{\text{VA}}^{L\mathcal{K}}(\tau,\tau-1) \tag{2.46}$$

$$\text{RVA}^{k}(t,b) \equiv \text{VA}^{kb} \prod_{\tau=b+1}^{t} Q_{\text{VA}}^{Lk}(\tau,\tau-1)\ (k\in\mathcal{K}), \tag{2.47}$$

where the superscript L denotes Laspeyres. Nominal reference period values are uprated by chained Laspeyres quantity indices. This system is clearly non-additive. The extent of non-additivity depends of course on relative price developments.

For those who want to overcome non-additivity, Balk and Reich (2008) proposed the following set of deflators:

$$P_{\text{VA}}^{\mathcal{K}}(t,b) \equiv \prod_{\tau=b+1}^{t} P_{\text{VA}}^{P\mathcal{K}}(\tau,\tau-1) \tag{2.48}$$

$$P_{\text{VA}}^{k}(t,b) \equiv P_{\text{VA}}^{Pk}(t,t-1) \prod_{\tau=b+1}^{t-1} P_{\text{VA}}^{P\mathcal{K}}(\tau,\tau-1)\ (k\in\mathcal{K}). \tag{2.49}$$

Notice the subtle difference between the expressions (2.45) and (2.49): in the Balk-Reich approach, the k-specific deflators differ only in the $(t, t-1)$ stretch, whereas the tails, covering the $(t-1, b)$ stretch, are the same. It is straightforward to verify that this system returns additive real values. Recall that the definition of the Paasche price index implies the following identity,

$$\frac{\text{VA}^{\mathcal{K}t}}{P_{\text{VA}}^{P\mathcal{K}}(t,t-1)} = \sum_{k\in\mathcal{K}} \frac{\text{VA}^{kt}}{P_{\text{VA}}^{Pk}(t,t-1)}. \tag{2.50}$$

Dividing both sides of equation (2.50) by $\prod_{\tau=b+1}^{t-1} P_{VA}^{P\mathcal{K}}(\tau, \tau-1)$, and applying the definition of real value added, then delivers

$$\mathrm{RVA}^{\mathcal{K}}(t,b) = \sum_{k\in\mathcal{K}} \mathrm{RVA}^{k}(t,b), \quad (2.51)$$

which means additivity.

Recently, Choi (2015) proposed an alternative. To see how this works, notice that equation (2.50) is equivalent to

$$Q_{VA}^{L\mathcal{K}}(t,t-1) = \sum_{k\in\mathcal{K}} \frac{\mathrm{VA}^{k,t-1}}{\mathrm{VA}^{\mathcal{K},t-1}} Q_{VA}^{Lk}(t,t-1). \quad (2.52)$$

Now, Choi's Condition of Internal Consistency (CIC) was defined as

$$\frac{\mathrm{RVA}^{k}(t,b) - \mathrm{RVA}^{k}(t-1,b)}{\mathrm{RVA}^{\mathcal{K}}(t-1,b)} = \frac{\mathrm{VA}^{k,t-1}}{\mathrm{VA}^{\mathcal{K},t-1}}\left(Q_{VA}^{Lk}(t,t-1) - 1\right) \; (k \in \mathcal{K}). \quad (2.53)$$

On the right-hand side of this equation, we see the contribution of production unit k to aggregate quantity change $Q_{VA}^{L\mathcal{K}}(t,t-1) - 1$. On the left-hand side, we see the contribution of the same production unit to aggregate real value added change $\mathrm{RVA}^{\mathcal{K}}(t,b)/\mathrm{RVA}^{\mathcal{K}}(t-1,b) - 1$, if real value added were additive.

Expression (2.53) can be rewritten as

$$\begin{aligned}\mathrm{RVA}^{k}(t,b) &= \mathrm{RVA}^{k}(t-1,b) \\ &\quad + \mathrm{RVA}^{\mathcal{K}}(t-1,b)\frac{\mathrm{VA}^{k,t-1}}{\mathrm{VA}^{\mathcal{K},t-1}}\left(Q_{VA}^{Lk}(t,t-1) - 1\right) \; (k \in \mathcal{K}).\end{aligned} \quad (2.54)$$

Summing over all the production units and using expression (2.52) delivers the following result,

$$\begin{aligned}\sum_{k\in\mathcal{K}} \mathrm{RVA}^{k}(t,b) &= \sum_{k\in\mathcal{K}} \mathrm{RVA}^{k}(t-1,b) + \mathrm{RVA}^{\mathcal{K}}(t-1,b)\left(Q_{VA}^{L\mathcal{K}}(t,t-1) - 1\right) \\ &= \sum_{k\in\mathcal{K}} \mathrm{RVA}^{k}(t-1,b) + \mathrm{RVA}^{\mathcal{K}}(t,b) - \mathrm{RVA}^{\mathcal{K}}(t-1,b),\end{aligned} \quad (2.55)$$

where the final step rests on expression (2.46). The first line and last line of expression (2.55) taken together mean that if additivity holds in period

$t-1$, then also in period t; formally,

$$\mathrm{RVA}^{\mathcal{K}}(t-1,b)=\sum_{k\in\mathcal{K}}\mathrm{RVA}^{k}(t-1,b)\Rightarrow\mathrm{RVA}^{\mathcal{K}}(t,b)=\sum_{k\in\mathcal{K}}\mathrm{RVA}^{k}(t,b). \tag{2.56}$$

Since additivity obviously holds in the reference period b, $\mathrm{RVA}^{\mathcal{K}}(b,b)=\sum_{k\in\mathcal{K}}\mathrm{RVA}^{k}(b,b)$, we may conclude that Choi's CIC generates an additive system of aggregate and sub-aggregate value added.

A disadvantage of Choi's system is that it does not provide explicit functional forms for the aggregate and sub-aggregate deflators. Instead, deflators are defined as ratios of nominal to real value added.

Conclusion

The two key results of this chapter are expression (2.17), concerning the decomposition of aggregate output (= real value added) growth, and expression (2.24), concerning the decomposition of labour productivity growth. If additivity is assumed, then aggregate labour productivity appears to take on a very simple form. Against this backdrop a number of misallocation measures were reviewed. The final section was devoted to a comparison of two ways of enforcing additivity on non-additive data.

References

Baldwin, J. R., & Willox, M. (2016). The industry origins of Canada's weaker labour productivity performance and the role of structural adjustment in the post-2000 period. *International Productivity Monitor, 31*, 19–36.

Balk, B. M. (2008). *Price and quantity index numbers: Models for measuring aggregate change and difference*. Cambridge University Press.

Balk, B. M. (2014). Dissecting aggregate output and labour productivity change. *Journal of Productivity Analysis, 42*, 35–43.

Balk, B. M. (2021). *Productivity: concepts, measurement, aggregation, and decomposition*. Contributions to Economics: Springer Nature Switzerland AG.

Balk, B. M., & Reich, U. P. (2008). Additivity of national accounts reconsidered. *Journal of Economic and Social Measurement, 33*, 165–178.

Bartelsman, E. J., & Wolf, Z. (2018). Measuring Productivity Dispersion. In E. Grifell-Tatjé, C. A. K. Lovell, & R. C. Sickles (Eds.), *The Oxford Handbook of Productivity Analysis*. New York: Oxford University Press.

Berlingieri, G., Blanchenay, P., Calligaris, S., & Criscuolo, C. (2017). *The multiprod project: A comprehensive overview* (OECD Science, Technology and Industry Working Papers 2017/04). Organisation for Economic Co-operation and Development, Paris .

Choi, K. H. (2015). *How can we restore the additivity of chained national accounts?* Paper in progress, personal communication.

De Loecker, J., & Syverson, C. (2021). *An industrial organization perspective on productivity* (Working Paper 29229). National Bureau of Economic Research, Cambridge MA. Forthcoming in Handbook of Industrial Organization, Volume 4. Elsevier.

Diewert, W. E. (2010). On the Tang and Wang decomposition of labour productivity growth into sectoral effects. In W. E. Diewert, B. M. Balk, D. Fixler, K. J. Fox, & A. O. Nakamura (Eds.), *Price and productivity measurement: Volume 6—Index number theory*. Trafford Press, www.vancouvervolumes.com and www.indexmeasures.com

Diewert, W. E. (2015). Decompositions of productivity growth into sectoral effects. *Journal of Productivity Analysis, 43*, 367–387.

Diewert, W. E. (2016). Decompositions of productivity growth into sectoral effects: Some puzzles explained. In W. H. Greene, L. Khalaf, R. C. Sickles, M. Veall, & M.-C. Voia (Eds.), *Productivity and efficiency analysis*. Proceedings in Business and Economics. Springer International Publishing.

Dumagan, J. C. (2013a). A generalized exactly additive decomposition of aggregate labor productivity growth. *The Review of Income and Wealth, 59*, 157–168.

Dumagan, J. C. (2013b). *Relative price effects on decompositions of change in aggregate labor productivity* (Discussion Paper Series No. 2013-44). Philippine Institute for Development Studies, Makati City.

Dumagan, J. C. (2014). *Consistent level aggregation and growth decomposition of real GDP* (Working Paper Series 2014-008). Angelo King Institute for Economic and Business Studies, De La Salle University, Manila, Philippines.

Dumagan, J. C. (2016). *Effects of relative prices on contributions to the level and growth of real GDP* (Working Paper Series 2016-036). Angelo King Institute for Economic and Business Studies, De La Salle University, Manila, Philippines.

Dumagan, J. C., & Balk, B. M. (2016). Dissecting aggregate output and labour productivity change: A postscript on the role of relative prices. *Journal of Productivity Analysis, 45*, 117–119.

Foster, L., Grim, C., Haltiwanger, J. C., & Wolf, Z. (2021). Innovation, productivity dispersion, and productivity growth. In C. Corrado, J. Haskel, J. Miranda, & D. Sichel (Eds.), *Measuring and accounting for innovation in the twenty-first century*. NBER Studies in Income and Wealth, Volume 78. The University of Chicago Press.

Gollop, F. M. (1979). Accounting for intermediate input: The link between sectoral and aggregate measures of productivity growth. In A. Rees (chairman), Measurement and Interpretation of Productivity. Report of the Panel to Review Productivity Statistics, Committee on National Statistics, Assembly of Behavioral and Social Sciences, National Research Council. National Academy of Sciences, Washington DC.

Gu, W. (2019). Frontier firms, productivity dispersion and aggregate productivity growth in Canada. *International Productivity Monitor, 37*, 96–119.

Haltiwanger, J., Kulick, R., & Syverson, C. (2018). *Misallocation measures: The distortion that ate the residual* (Working Paper 24199). National Bureau of Economic Research, Cambridge MA.

Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics, 6*, 735–770.

Hsieh, C. T., & Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics, 124*, 1403–1448.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *Review of Economic Studies, 45*, 511–518.

Jorgenson, D. W., Gollop, F. M., & Fraumeni, B. M. (1987). *Productivity and U. S. Economic Growth*. Harvard University Press.

Nordhaus, W. D. (2002). Productivity growth and the new economy. *Brookings Papers on Economic Activity,* (2), 211–244.

Reinsdorf, M., & Yuskavage, R. (2010). Exact industry contributions to labor productivity change. In W. E. Diewert, B. M. Balk, D. Fixler, K. J. Fox, & A. O. Nakamura (Eds.), *Price and productivity measurement: Volume 6— Index number theory*. Trafford Press, www.vancouvervolumes.com and www.indexmeasures.com.

Stiroh, K. J. (2002). Information technology and the U.S. productivity revival: What do the industry data say? *The American Economic Review, 92*, 1559–1576.

Tang, J., & Wang, W. (2004). Sources of aggregate labour productivity growth in Canada and the United States. *Canadian Journal of Economics, 37*, 421–444.

Tang, J., & Wang, W. (2015). Unbalanced industry demand and supply shifts: Implications for economic growth in Canada and the United States. *The Review of Income and Wealth, 61*, 773–798.

CHAPTER 3

Efficiency Analysis with Stochastic Frontier Models Using Popular Statistical Softwares

Bao Hoang Nguyen, Robin C. Sickles, and Valentin Zelenyuk

Introduction

In this chapter, we provide practitioners, who are interested in analysing the performance of production units, with a brief introduction to the stochastic frontier paradigm—one of the most powerful techniques for

B. H. Nguyen
School of Economics, The University of Queensland, Brisbane, QLD, Australia
e-mail: bao.nguyen3@uq.net.au

R. C. Sickles
Economics Department, Rice University, Houston, TX, USA
e-mail: rsickles@rice.edu

V. Zelenyuk (✉)
School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, Brisbane, QLD, Australia
e-mail: v.zelenyuk@uq.edu.au

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_3

performance analysis developed in the last century.[1] Stochastic frontier analysis employs econometric models to estimate production frontiers and technical (in)efficiency with respect to these frontiers. Since its first introduction by Aigner et al. (1977) and Meeusen and van den Broeck (1977), stochastic frontier analysis has been applied to study the productivity and efficiency of production units in various economic sectors, such as banking (e.g., Adams et al., 1999; Ferrier & Lovell, 1990; Kumbhakar & Tsionas, 2005; Malikov et al., 2016), healthcare (e.g., Comans et al., 2020; Greene, 2004; Mutter et al., 2013; Rosko, 2001; Zuckerman et al., 1994), and agriculture (e.g., Battese & Broca, 1997; Battese & Coelli, 1995; Kumbhakar & Tsionas, 2008), to mention a few. Moreover, the methodology is also used to undertake cross-country studies on various important aspects of society such as the healthcare system (Greene, 2004) and taxation (Fenochietto & Pessino, 2013).

Our chapter also documents the estimation routines used to implement the classical models as well as the recent developments in this research area for practitioners, especially those who are willing to use Stata, but also with tips on where to find analogous programs for R and Matlab users.[2] Interested readers can find more comprehensive overviews in Sickles and Zelenyuk (2019, Chap. 11–16) and Kumbhakar et al. (2021a, 2021b).

The structure of this chapter is as follows. We start our discussion with the basic stochastic frontier model. We then extend our discussion to various generalisations of the stochastic frontier paradigm, including stochastic panel data models, stochastic frontier models with determinants of inefficiency, also referred to in the literature as "environmental factors", and the semi-parametric stochastic frontier models. To provide readers with an accessible toolkit to implement these methods, we also document available commands/packages in popular statistical softwares. We focus

[1] Another powerful technique for performance analysis is data envelopment analysis—the technique based on the mathematical linear programming method proposed by Farrell (1957) and popularised by Charnes et al. (1978).

[2] On this aspect, our chapter complements earlier surveys on empirical frontier application and productivity and efficiency analysis software, e.g., Daraio et al. (2019, 2020). Besides, the chapter also complements the previous contributions of Belotti et al. (2013) and Kumbhakar et al. (2015), who focused only on stochastic frontier analysis using Stata, by providing the sources on analogous implementations in Matlab and R. Moreover, we also include the discussion about the semi-parametric stochastic frontier models with ready-to-use Stata codes to implement the model proposed by Simar et al. (2017), which to the best of our knowledge have not been documented elsewhere before.

on the implementation via Stata and also provide brief comments on the sources on analogous implementations in Matlab and R. Besides, we also provide an empirical illustration for the methods discussed in this chapter.

Basic Stochastic Frontier Models

The stochastic frontier paradigm can be viewed as a generalisation of the classical production function approach, where the optimal allocation in production is a testable restriction rather than a prior assumption usually assumed by the neoclassical production theory (Sickles & Zelenyuk, 2019).

The distinctive feature of the stochastic frontier paradigm (compared to the canonical average production model paradigm) is its non-symmetric two-component error, composed of a regular idiosyncratic disturbance and an additional one-sided non-negative error component.[3] The former accounts for factors such as measurement error, misspecification, and the randomness of the production process, whereas the latter aims to represent the technical inefficiency that reduces the actual output from its maximum feasible level.[4] Assumptions in the canonical model used in stochastic frontier analysis on the conditional independence of both error terms and the regressors as well as their independence from each other have been lifted over the years in a series of refinements of the basic model. We will discuss these in turn later in our chapter.

Aigner et al. (1977) Model

The canonical model of the stochastic frontier paradigm was proposed independently by Aigner et al. (1977) (hereafter ALS) and Meeusen and

[3] In the panel data context, which we will discuss in the next sections, the composed error can include four components.

[4] In this chapter, our discussion will follow the traditional exposition based on the production function. A similar exposition (with some adaptations) applies to other characterisations of the production side, such as cost function and revenue function. Meanwhile, more elaboration is needed if one is interested in measuring profit efficiency (see Färe et al., 2019; Sickles & Zelenyuk, 2019 Chap. 2) and references therein.

van den Broeck (1977). The ALS model is formulated as[5]

$$\begin{aligned}
\ln y_i &= \ln f(x_i|\beta) + \varepsilon_i, \quad i = 1, \ldots, n, \\
\varepsilon_i &= v_i - u_i, \\
v_i &\sim_{\text{iid}} \mathcal{N}\left(0, \sigma_v^2\right), \\
u_i &\sim_{\text{iid}} \mathcal{N}^+\left(0, \sigma_u^2\right),
\end{aligned} \tag{3.1}$$

where $y_i \in \Re_+^1$ is the output, $x_i \in \Re_+^p$ is a vector of p inputs and β is a vector of the parameters corresponding to x_i.[6] The error term ε_i is composed of a normally distributed disturbance, v_i, representing the measurement and specification error, and a positive disturbance u_i (following the half-normal distribution), representing technical inefficiency.[7] Furthermore, v_i and u_i are assumed to be statistically independent from each other and from x_i. With the distribution assumptions on u_i and v_i, the likelihood function for the model is constructed and the model is then estimated using the maximum likelihood estimator.

Once the parameters of the model have been estimated, one can obtain the expected level of technical inefficiency by estimating

$$E[u_i] = \sqrt{2/\pi}\sigma_u, \tag{3.2}$$

[5] The formulation here is a convenient representation of a production relationship, where actual output is decomposed into the maximum output (with noise) and inefficiency, i.e., $y_i = f(x_i|\beta)\exp(\varepsilon_i) = f(x_i|\beta)\exp(v_i)\exp(-u_i)$. After log-transformation, we have a linear relationship as shown in Eq. (3.1).

[6] Multiple outputs also can be considered. For example, this can be done by employing a distance function instead of the production function or by looking at the estimation of the cost frontier or by converting outputs into polar coordinates (e.g., see Simar & Zelenyuk, 2011). One can also use dimension reduction techniques to reduce the dimension outputs or inputs into smaller dimensions, e.g., via Principle Component Analysis, or using economic or price-based aggregation (e.g., see related discussion in Zelenyuk (2020) and an application in Nguyen and Zelenyuk [2021]). The latter approach can be especially useful in the case of very large dimensions (sometimes called 'big wide data' cases), e.g., as is done for measuring the total output of countries (e.g., GDP), industries or firms (total revenue) or for some inputs (e.g., capital). Due to space limitation, we will focus here on the single output case, as was also considered in ALS and many other studies.

[7] Other distributional assumptions such as exponential, truncated normal, gamma, and so on, can be used for the inefficiency term (e.g., see Almanidis & Sickles, 2012; Almanidis et al., 2014; Greene, 1980a, 1980b, 1990; Meeusen & van den Broeck, 1977; Stevenson, 1980).

and the expected level of efficiency by using the following approximation[8]

$$E\left[\exp(-u_i)\right] \approx 1 - E[u_i]. \tag{3.3}$$

If one is interested in the estimates of individual (in)efficiency of a specific production unit, more elaboration is needed. The most popular approach in the literature is to follow Jondrow et al. (1982) (hereafter JLMS), where the inefficiency of a production unit can be estimated or predicted using the expected value of u_i conditional on the realisation of the composed error of the model, i.e., $E(u_i|\varepsilon_i)$,[9] given by

$$E(u_i|\varepsilon_i) = \frac{\sigma_* \phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)} + \mu_{*i}, \tag{3.4}$$

where

$$\mu_{*i} = \frac{-\sigma_u^2 \varepsilon_i}{\sigma_v^2 + \sigma_u^2}, \tag{3.5}$$

and

$$\sigma_*^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}, \tag{3.6}$$

while $\phi(\cdot)$ and $\Phi(\cdot)$ are *pdf* and *cdf* of the standard normal distribution, respectively.[10] It is worth noting that while being originally developed for ALS, the JLMS-type procedure can be extended to predict (in)efficiency of a specific firm in the other models estimated by the maximum likelihood estimator (see more discussion in Kumbhakar, 1987).

[8] The exact expression of the expected level of efficiency is given by $E\left[\exp(-u_i)\right] = 2\Phi(-\sigma_u)\exp\left(\frac{\sigma_u^2}{2}\right)$.

[9] It is worth noting here that although this estimator is unbiased, it is an inconsistent estimator of individual inefficiency (see more discussion in Jondrow et al., 1982).

[10] One also can estimate the efficiency of a production unit by using the relationship $E\left[\exp(-u_i)|\varepsilon_i\right] \approx 1 - E\left[u_i|\varepsilon_i\right]$ or utilising the exact expression $E\left[\exp(-u_i)|\varepsilon_i\right] = \exp\left(-\mu_{*i} + \frac{1}{2}\sigma_*^2\right)\frac{\Phi\left(\frac{\mu_{*i}}{\sigma_*} - \sigma_*\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}$ (Battese & Coelli, 1988).

Implementation of ALS Model

There are several options to estimate the basic stochastic frontier model in Stata. One can use the official Stata command `frontier` or utilise the command `sfcross` written by Belotti et al. (2013)[11] or even set up the likelihood function using the `sfmodel` command then estimating the model using the official Stata routine for the maximum likelihood, `ml max`, as described in the handbook of Kumbhakar et al. (2015).[12] These commands generate equivalent results for the basic stochastic frontier models and differ only in the formatting and listing of results and the options available for the different treatments of error distributions for the one-sided efficiency term and the inclusion of environmental factors.

As we progress in our chapter, we consider a richer set of generalisations of the canonical stochastic frontier paradigm. Also, user-written commands provide us with more flexibility to estimate models that are not available with the current official Stata commands. Moreover, the user-written commands by Belotti et al. (2013) and Kumbhakar et al. (2015) also equip us with options to provide and refine the initial values for the maximum likelihood estimation, which can be very useful when dealing with complex likelihood functions.

After estimating the models, the estimates of technical inefficiency and efficiency can be obtained by using the postestimation routine `predict` (for the models estimated in the Stata version 16 by the official Stata command and the command written by Belotti et al. [2013]) or `sf_predict` (for the models estimated by the command written by Kumbhakar et al. [2015]). As an illustration, a snippet of Stata codes for implementing the ALS model is provided in Box 3.1.

R software also has several packages to implement the estimation of the basic stochastic frontier model. For example, one can use the package

[11] The `sfcross` command (and the `sfpanel` command that we will discuss later for the panel data context) can be installed by executing the following command lines in Stata: ssc install sfcross and ssc install sfpanel.

[12] The `sfmodel` and other user-written commands provided in the handbook of Kumbhakar et al. (2015) can be installed in Stata by executing the following command lines: `net install sfbook_install, from` (https://sites.google.com/site/sfbook2014/home/install/) `replace` and `sfbook_install` (see more details in Kumbhakar et al. 2015 and its website, https://sites.google.com/site/sfbook2014/).

Box 3.1 Illustration for the implementation of the Aigner et al. (1977) model

```
************************************************************************
* Illustration for the implementation of the Aigner et al. (1977) model
***************************Partial Stata Codes**************************
************************************************************************
/* Note that output and inputs are in log forms and stored in global
   Stata variables $y and $xlist, respectively */
/* Implementation using the standard Stata commands */
frontier $y $xlist, distribution(hnormal)
predict ineff_ALS_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_ALS_1, te /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Belotti et al. (2013) */
sfcross $y $xlist, distribution(hnormal)
predict ineff_ALS_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_ALS_2, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Kumbhakar et al. (2015) */
sfmodel $y, prod frontier($xlist) distribution(h)
ml max
sf_predict, jlms(ineff_ALS_3) /* Predict inefficiency, i.e., E(u|e) */
sf_predict, bc(eff_ALS_3) /* Predict efficiency, i.e., E(exp(-u)|e) */
```

`frontier` written by Coelli and Henningsen (2020)[13] or utilise the function `sfa` in the package `Benchmarking` written by Bogetoft and Otto (2019).

In order to estimate the basic stochastic frontier model, Matlab users need to set up the likelihood function and then utilise the optimisation routines, such as `fminunc` to optimise the likelihood function. Sickles and Zelenyuk (2019) provide a suite of Matlab codes to estimate a variety of stochastic frontier models on the website that accompanies their book.[14,15] Although they do not include the ALS model, one can easily

[13] The package `frontier` uses the Fortran source codes of Frontier 4.1 originally developed by Tim Coelli (see more details in the manual of the package available at https://cran.r-project.org/web/packages/frontier/frontier.pdf).

[14] The website can be found at https://sites.google.com/site/productivityefficiency/home.

[15] The Matlab codes accompanying Sickles and Zelenyuk (2019) are also converted to R codes by Sickles et al. (2020), which can be accessed via the link provided on the book website or directly via https://sites.google.com/site/productivityinr.

adapt their codes to obtain the estimates for this basic stochastic frontier model.

Early Generation of Stochastic Panel Data Models

The basic stochastic frontier model discussed in the previous section is formulated in the cross-sectional setting and suffers from a number of drawbacks. As discussed in Schmidt and Sickles (1984), the three main disadvantages of the basic cross-sectional stochastic frontier model are: (i) there does not exist a consistent estimator of individual efficiency, (ii) the parametric distributional assumptions are usually required for the two error components (inefficiency and noise) to estimate the model and to predict the overall and individual (in)efficiency, and (iii) the assumption that inefficiency is independent of regressors is usually not plausible.

Over the past four decades, substantial efforts have been made to address these drawbacks of the cross-sectional stochastic frontier model. Among those, particular interest hinges on exploiting the advantages of panel data structure. Schmidt and Sickles (1984) were among the first who provided a general framework to extend the cross-sectional stochastic frontier model to the panel data setting, which also encompasses the Pitt and Lee (1981) full parametric random effects model.

Schmidt and Sickles (1984) Model

The model in Schmidt and Sickles (1984) can be formulated as follows

$$y_{it} = \beta_0 + x'_{it}\beta + v_{it} - u_i, \quad i = 1, \ldots, n, \; t = 1, \ldots, T, \tag{3.7}$$

where $y_{it} \in \Re^1_+$ is the output, $x_{it} \in \Re^p_+$ is a vector of p inputs of production unit i in time t. v_{it} is the regular disturbance, while the unobserved individual heterogeneity, u_i, represents technical inefficiency. Model (3.7) can be rewritten as

$$y_{it} = \beta_0^* + x'_{it}\beta + v_{it} - u_i^* = c_i + x'_{it}\beta + v_{it}, \tag{3.8}$$

where $\beta_0^* = \beta_0 - E(u_i)$, $u_i^* = u_i - E(u_i)$, $E(u_i) \geq 0$, $c_i = \beta_0^* - u_i^* = \beta_0 - u_i$.

Model (3.8) turns out to be a usual panel data model and can be estimated using the standard estimation methods in the panel data literature, such as the within estimator (i.e., in the fixed effects framework), the generalised least-square estimator (i.e., in the random effects framework), and the Hausman-Taylor estimator. After estimating the model, one can obtain the estimate $\hat{c}_i$ of c_i and follow Schmidt and Sickles (1984) to construct a consistent estimator of technical inefficiency

$$\hat{u}_i = \max(\hat{c}_i) - \hat{c}_i \geq 0, \quad i = 1, \ldots, n. \tag{3.9}$$

The estimated inefficiency in Eq. (3.9) is measured with respect to the best practice production unit in the sample, which is implicitly assumed to be 100% efficient.

Implementation of Schmidt and Sickles (1984) Model

One can estimate the Schmidt and Sickles (1984) model using standard routines in Stata. Specifically, the official Stata command `xtreg` can be utilised to estimate the standard panel data model in Eq. (3.8) and the postestimation command `predict` can be used to obtain the estimate $\hat{c}_i$ of c_i. It is then straightforward to code formula (3.9) into Stata to get the estimates of technical inefficiency. Alternatively, one can use the command `sfpanel` written by Belotti et al. (2013) with the option `model(fe)` or `model(regls)` to estimate Schmidt and Sickles (1984) model in a fixed or random effects framework, respectively. As an illustration, a snippet of Stata codes for implementing the Schmidt and Sickles (1984) model (in the fixed effects framework) is provided in Box 3.2.

It is worth noting that model (3.8) and the individual inefficiency in (3.9) are estimated without any parametric assumptions on the distributions of composed errors. Alternatively, one can impose parametric assumptions on the distributions of the error components in model (3.7), e.g., a half-normal distribution for u_i and a normal distribution for v_{it} as discussed in Pitt and Lee (1981) and Schmidt and Sickles (1984). The model then can be estimated using the maximum likelihood estimator and the individual technical efficiency can be obtained by employing the JLMS procedure (extended to the panel data setting by Kumbhakar, 1987). This model is estimated in Stata using the user-written command `sfpanel` from Belotti et al. (2013) with the option `model(pl81)`. Alternatively, if one assumes that u_i follows a truncated normal distribution, i.e., $u_i \sim \mathcal{N}^+(\mu, \sigma_u^2)$, then the official Stata command `xtfrontier` with the option `ti` can be utilised. A snippet of Stata codes for implementing the Pitt and Lee (1981) model is provided in Box 3.3.

Box 3.2 Illustration for the implementation of the Schmidt and Sickles (1984) model in the fixed effects framework

```
*********************************************************************
*********************Illustration for the implementation *************
*******************of the Schmidt and Sickles (1984) model ************
***************************Partial Stata Codes************************
*********************************************************************
/* The illustration here is for the fixed effects framework */
/* Note that output and inputs are in log forms and stored in global
   Stata variables $y and $xlist, respectively */
/* Implementation using the standard Stata commands */
xtreg $y $xlist, fe /* Need to declare data to be panel before using
   xtreg command*/
predict ci, u /* Obtain the estimate of ci */
quietly summarize ci
gen ineff_SS_1 = r(max) - ci /* Predict inefficiency*/
gen eff_SS_1 = exp(-ineff_SS_1) /* Predict efficiency*/
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(fe)
predict ineff_SS_2, u /* Predict inefficiency*/
gen eff_SS_2 = exp(-ineff_SS_2) /* Predict efficiency*/
```

Box 3.3 Illustration for the implementation of the Pitt and Lee (1981) model in the random effects framework

```
*********************************************************************
* Illustration for the implementation of the Pitt and Lee (1981) model*
***************************Partial Stata Codes************************
*********************************************************************
/* Note that output and inputs are in log forms and stored in global
   Stata variables $y and $xlist, respectively */
/* Implementation using the standard Stata commands */
xtfrontier $y $xlist, ti
predict ineff_PL_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_PL_1, te /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(pl81)
predict ineff_PL_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_PL_2, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
```

Estimation of the Schmidt and Sickles (1984) model also can be implemented in Matlab and R using the codes provided by Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the links in footnotes 14 and 15).

Cornwell et al. (1990) Model

The technical inefficiency estimated within the Schmidt and Sickles (1984) framework is time-invariant, which may be an unrealistic restriction in many applied settings, especially in a long panel. To allow for time-varying inefficiency in the Schmidt and Sickles (1984) framework, one can follow the suggestion in Cornwell et al. (1990) to replace c_i by, e.g., c_{it}, where c_{it} is a quadratic function of time trend t with the parameters (coefficients) being firm-specific, in particular

$$c_{it} = \theta_{0i} + \theta_{1i}t + \theta_{2i}t^2. \tag{3.10}$$

The parameters in Eq. (3.10) can be estimated by regressing the residual from the model (3.7) for production unit i on a constant, time, and time-squared (see more discussion in Cornwell et al., 1990). The fitted value from this model provides us with a consistent estimate (for large N) of c_{it}, denoted as $\hat{c}_{it}$. The individual technical inefficiency of production unit i at time t then can be estimated using an analogous procedure to Schmidt and Sickles (1984), specifically[16]

$$\hat{u}_{it} = \hat{c}_t - \hat{c}_{it}, \tag{3.11}$$

where

$$\hat{c}_t = \max_j \left(\hat{c}_{jt}\right), \quad t = 1, \ldots, T. \tag{3.12}$$

[16] Cornwell et al. (1990) outlined estimators for a general model in which any set of regressors could be drivers of efficiency change, if efficiency was interpreted as firm-specific heterogeneity. These regressors could be time varying. Thus, the Cornwell et al. (1990) model was the first study about which we are aware to address the issue of environmental variables influencing efficiency levels.

Box 3.4 Illustration for the implementation of the Cornwell et al. (1990) model

```
**********************************************************************
*Illustration for the implementation of the Cornwell et al. (1990) model
**************************Partial Stata Codes************************
**********************************************************************
/* Note that output and inputs are in log forms and stored in global
   Stata variables $y and $xlist, respectively */
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(fecss)
predict ineff_CSS, u /* Predict inefficiency*/
gen eff_CSS = exp(-ineff_CSS) /* Predict efficiency*/
```

Implementation of Cornwell et al. (1990) Model

Estimation of the Cornwell et al. (1990) model can be implemented using standard Stata routines in a set of procedures similar to those we discussed for the Schmidt and Sickles (1984) model. Alternatively, one can utilise the user-written command `sfpanel` from Belotti et al. (2013) with the option `model(fecss)`. A snippet of codes for implementing the Cornwell et al. (1990) model using the `sfpanel` command is provided in Box 3.4.

Being similar to the Schmidt and Sickles (1984) model, one can estimate the Cornwell et al. (1990) model in Matlab and R using the codes provided by Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the links in footnotes 14 and 15).

Kumbhakar (1990) and Battese and Coelli (1992) Models

If one is willing to impose distributional assumptions on the inefficiency component (as well as on the random disturbance term), the maximum likelihood estimation can be utilised to estimate time-varying efficiency models. Kumbhakar (1990) and Battese and Coelli (1992) appear to be the most popular models of this type. In the Kumbhakar (1990) model, time-varying inefficiency is modelled as

$$\begin{aligned} u_{it} &= \left(1 + \exp\left(at + bt^2\right)\right)^{-1} \tau_i, \\ \tau_i &\sim_{\text{iid}} \mathcal{N}^+\left(0, \sigma_\tau^2\right), \end{aligned} \tag{3.13}$$

while in Battese and Coelli (1992), time-varying inefficiency is specified as

$$
\begin{aligned}
u_{it} &= \{\exp[-\eta(t - T)]\}\tau_i, \\
\tau_i &\sim_{\text{iid}} \mathcal{N}^+\left(\mu, \sigma_\tau^2\right),
\end{aligned}
\tag{3.14}
$$

where a, b, and η are parameters to be estimated, and in both models, the random disturbance follows a normal distribution, i.e., $v_{it} \sim_{\text{iid}} \mathcal{N}\left(0, \sigma_v^2\right)$.

Being similar to the Cornwell et al. (1990) model, the Kumbhakar (1990) and Battese and Coelli (1992) models extend the Pitt and Lee (1981) model by allowing the mean of inefficiency to vary over time, but they are more parsimonious in the sense that temporal patterns only depend on one or two parameters. The Cornwell et al. (1990) model, however, has an advantage in that it allows temporal patterns to vary across production units. Moreover, as discussed above, estimation of the Cornwell et al. (1990) model does not require parametric assumptions for the inefficiency term.

Implementation of Kumbhakar (1990) and Battese and Coelli (1992) Models

The Battese and Coelli (1992) model, also known as a "time decay" model, can be estimated using Stata commands in its version 16 platform as well as by using user-written commands. Specifically, the estimation can be implemented by using the `xtfrontier` command with the option `tvd` or the command `sfpanel` from Belotti et al. (2013) with the option `model(bc92)`. The official Stata command `xtfrontier` cannot carry out the estimation of the Kumbhakar (1990) model, which is available using the option `model(kumb90)` with the command `sfpanel` from Belotti et al. (2013). A snippet of Stata codes for implementing Kumbhakar (1990) and Battese and Coelli (1992) models is provided in Box 3.5.

The estimation of the Battese and Coelli (1992) model can be implemented in R software by using the package `frontier` written by Coelli and Henningsen (2020). Alternatively, R users and Matlab users can utilise the codes prepared by Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the links in footnotes 14 and 15).

Box 3.5 Illustration for the implementation of the Kumbhakar (1990) and Battese and Coelli (1992) models

```
***********************************************************************
***** Illustration for the implementation of the Kumbhakar (1990)  *****
************* and Battese and Coelli (1992) models *******************
*********************** Partial Stata Codes **************************
***********************************************************************
/* Note that output and inputs are in log forms and stored in global
   Stata variables $y and $xlist, respectively */
/* Implementation using the standard Stata commands */
xtfrontier $y $xlist, tvd /* the Battese and Coelli (1992) model */
predict ineff_BC_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_BC_1, te /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(bc92) /* the Battese and Coelli (1992) model */
predict ineff_BC_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_BC_2, bc /* Predict efficiency, i.e., E(exp(-u)|e) */

sfpanel $y $xlist, model(kumb90) /* the Kumbhakar (1990) model */
predict ineff_K, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_K, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
```

Recent Advances of Stochastic Panel Data Models

The stochastic panel data models discussed so far have a major drawback in that technical inefficiency is not distinguishable from the unobserved individual heterogeneity, and thus, technical inefficiency confounds with all time-invariant unobserved individual effects. Various approaches have been proposed in the literature to mitigate this and other issues. Here, we will focus on a few, namely Greene (2005a, 2005b), Chen et al. (2014), Colombi et al. (2014), Kumbhakar et al. (2014), and Belotti and Ilardi (2018).

Greene (2005a, 2005b) Models

Greene (2005a, 2005b) proposed a stochastic panel data model in which unobserved individual heterogeneity separates from (transitory) technical efficiency. The model is formulated as

$$
\begin{aligned}
y_{it} &= c_i + x_{it}'\beta + \varepsilon_{it},\\
\varepsilon_{it} &= v_{it} - u_{it},
\end{aligned}
\tag{3.15}
$$

where v_{it} as before is a regular disturbance, while $u_{it} \geq 0$ is the source of inefficiency.

Estimation of the model in Eq. (3.15) is challenging, especially in the fixed effects framework. The two main challenges are: (i) the estimation of parameters may be inconsistent due to the incidental parameters problem, and (ii) there does not exist a closed-form expression of the likelihood function of the within or first-difference transformation of the composed error if one follows standard procedures. Greene (2005a) proposed to use the maximum likelihood dummy variable estimator to estimate the model in the fixed effects framework and provided simulation evidence showing that the incidental parameters problem is not serious for relatively large T.[17]

Implementation of Greene (2005a, 2005b) Models

One can implement the estimation of the Greene (2005a, 2005b) models in Stata by using the user-written command `sfpanel` from Belotti et al. (2013) with the option `model(tfe)` in the fixed effects framework and with the option `model(tre)` in the random effects framework.

Recently, Chen et al. (2014) derived a closed-form expression for the likelihood function of the within and first-difference transformation of the model by exploiting the properties of the closed-skew normal distribution class. The model in Eq. (3.15) then can be estimated consistently in the fixed effects framework using the marginal maximum likelihood estimator. Belotti and Ilardi (2018) further extend the work of Chen et al. (2014) by considering the simulated marginal maximum likelihood estimator.

Chen et al. (2014) and Belotti and Ilardi (2018) estimators can be implemented in Stata using the command `sftfe` written by Belotti and Ilardi (2018) with the options `estimator(within)` and

[17] Greene (2005a) also utilised the simulated maximum likelihood approach to estimate the model in the random effects framework.

Box 3.6 Illustration for the implementation of the Greene (2005a, 2005b) models

```
*************************************************************************
** Illustration for the implementation of the Green (2005a,b) models  **
****************************Partial Stata Codes*************************
*************************************************************************
/* The illustration here is for the fixed effect framework */
/* Note that output and inputs are in log forms and stored in global
   Stata variables $yand $xlist, respectively */
/* Implementation using  the maximum likelihood dummy variable estimator
    (the commands from Belotti et al. (2013)) */
sfpanel $y $xlist, distribution(hnormal) model(tfe)
predict ineff_G_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_G_1, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the marginal maximum likelihood estimator (the
   commands from Belotti and Ilardi (2018)) */
sftfe $y $xlist,  distribution(hnormal) estimator(within)
predict ineff_G_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_G_2, jlms /* Predict efficiency */
/* Implementation using the simulated marginal maximum likelihood
   estimator (the commands from Belotti and Ilardi (2018)) */
sftfe $y $xlist,  distribution(hnormal) estimator(mmsle)
predict ineff_G_3, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_G_3, jlms /* Predict efficiency */
```

`estimator(mmsle)`, respectively.[18] A snippet of Stata codes for implementing the Greene (2005a, 2005b) models is provided in Box 3.6.

To the best of our knowledge, routines to estimate the Greene (2005a, 2005b) model are not yet available in R and Matlab in a public domain.

Colombi et al. (2014) and Kumbhakar et al. (2014) Models

The model specified in Eq. (3.15), although distinguishing between unobserved individual heterogeneity and technical inefficiency, only considers transitory inefficiency. Kumbhakar et al. (2014) and Colombi et al. (2014) further extend the model by decomposing the inefficiency

[18] The `sftfe` command can be installed by executing the following command line in Stata: net install sftfe.pkg.

into transitory and persistent components whose formulation is

$$\begin{aligned} y_{it} &= \beta_0 + x_{it}'\beta + c_i - \eta_i + v_{it} - u_{it}, \\ c_i &\sim_{\text{iid}} \mathcal{N}\left(0, \sigma_c^2\right), \\ \eta_i &\sim_{\text{iid}} \mathcal{N}^+\left(0, \sigma_\eta^2\right), \\ v_{it} &\sim_{\text{iid}} \mathcal{N}\left(0, \sigma_v^2\right), \\ u_{it} &\sim_{\text{iid}} \mathcal{N}^+\left(0, \sigma_u^2\right), \end{aligned} \tag{3.16}$$

where c_i represents the unobserved individual heterogeneity, η_i represents the persistent inefficiency, u_{it} represents transitory inefficiency, and v_{it} is the regular disturbance. The model in Eq. (3.16) can be estimated using a single-stage maximum likelihood method (Colombi et al., 2014) or a multi-step procedure (Kumbhakar et al., 2014). The multi-step procedure although being inefficient relative to the single-stage maximum likelihood estimation, it is simpler and easier to implement. For the multi-step procedure, the model in Eq. (3.16) can be rewritten as

$$y_{it} = \beta_0^* + x_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{3.17}$$

where

$$\beta_0^* = \beta_0 - E[\eta_i] - E[u_{it}], \tag{3.18}$$

$$\alpha_i = c_i - \eta_i + E[\eta_i], \tag{3.19}$$

$$\varepsilon_{it} = v_{it} - u_{it} + E[u_{it}]. \tag{3.20}$$

The model in Eq. (3.17) turns out to be a standard panel data model and can be estimated by the usual panel data estimation methods. After estimating (3.17), one can obtain the predicted values of α_i and ε_{it}, $\hat{\alpha}_i$ and $\hat{\varepsilon}_{it}$, and then, the persistent and transitory inefficiency components are estimated by applying standard stochastic frontier techniques to (3.19) and (3.20) with α_i and ε_{it} replaced by $\hat{\alpha}_i$ and $\hat{\varepsilon}_{it}$, respectively.

Implementation of Colombi et al. (2014) and Kumbhakar et al. (2014) Models

The multi-step procedure to estimate the model specified in Colombi et al. (2014) and Kumbhakar et al. (2014) can be implemented in Stata using the command for panel data estimation, `xtreg`, together with the routines for basic stochastic frontier model estimation, e.g., `frontier` or `sfcross`. A snippet of Stata codes for implementing the Kumbhakar et al. (2014) model is provided in Box 3.7.

Similarly, R users can utilise panel data estimation routines (e.g., `plm` function) combined with commands for estimation of the basic stochastic frontier model discussed above (e.g., `sfa` or `frontier`) to estimate the Colombi et al. (2014) and Kumbhakar et al. (2014) models.

The implementation of the procedure in Matlab requires more effort since it is not easy (as in Stata or R) to perform panel data regression

Box 3.7 Illustration for the implementation of the Kumbhakar et al. (2014) model

```
************************************************************************
********************Illustration for the implementation ****************
**************** of the Kumbhakar et al. (2014) model ******************
****************************Partial Stata Codes*************************
************************************************************************
/* The illustration here is for the random effect framework */
/* Note that output and inputs are in log forms and stored in global
   Stata variables $y and $xlist, respectively */
/* Implementation using the standard Stata commands */
xtreg $y $xlist, re
predict alp, u /* Obtain estimates of alpha */
predict esl, e /* Obtain estimates of the composed error */
/* Estimate equation (19) using the basic stochastic frontier model to
   obtain persistent (in)efficiency */
frontier alp, distribution(hnormal)
predict ineff_pers, u /* Predict persistent inefficiency, E(u|e) */
predict eff_pers, te /* Predict persistent efficiency, E(exp(-u)|e) */
/* Estimate equation (19) using the basic stochastic frontier model to
   obtain transitory (in)efficiency */
frontier esl, distribution(hnormal)
predict ineff_trans, u /* Predict transitory inefficiency, E(u|e) */
predict eff_trans, te /* Predict transitory efficiency, E(exp(-u)|e) */
```

with this platform. With Matlab, one needs to write their own code or download and install the panel data toolbox, e.g., the one written by Álvarez et al. (2017), to estimate a panel data regression model.

Stochastic Frontier Models with Determinants of Inefficiency

An interesting generalisation of the stochastic frontier paradigm is extending the models to examine the impact of exogenous determinants on technical inefficiency. It is usually done by parameterising the parameters of inefficiency distribution, i.e., the pre-truncated mean and/or variance, as a function of exogenous variables. The approaches are applicable in both cross-sectional and panel data settings, and since it can be easily extended to panel data settings, here we focus our discussion on the cross-sectional context.

Popular Models

Cornwell et al. (1990) were the first to develop a model in which determinants of efficiency could be included in the stochastic frontier formulation. However, due to the linear way in which the determinants of efficiency were included in the regression model, their fixed effect estimator could not point identify both a covariate's effect on efficiency and its effect on the level of production. Kumbhakar et al. (1991) addressed this identification problem by specifying the efficiency determinants as a nonlinear function, parameterising the pre-truncated mean of inefficiency as a function of exogenous variable, specifically[19]

$$\begin{aligned} u_i &\sim \mathcal{N}^+\left(\mu_i, \sigma_u^2\right), \\ \mu_i &= z_i'\delta, \end{aligned} \tag{3.21}$$

where $z_i \in \Re^k$ is a vector of k exogenous variables (including the constant term) and δ is a vector of the parameters to be estimated. Alternatively, Caudill et al. (1995) proposed specifying the variance of the inefficiency

[19] This model specification was cast in the panel data context and popularised by Battese and Coelli (1995).

distribution as

$$
\begin{aligned}
u_i &\sim \mathcal{N}^+\left(0, \sigma_{ui}^2\right), \\
\sigma_{ui}^2 &= \exp\left(z_i'\delta\right).
\end{aligned}
\tag{3.22}
$$

One can also, at the same time, parameterise both the pre-truncated mean and variance of inefficiency as a function of the exogenous variables, i.e., combining (3.21) and (3.22), as in Wang (2002). These parametric stochastic frontier models are typically estimated using the maximum likelihood estimator in much the same way as the basic stochastic frontier model.

Wang and Schmidt (2002) suggested a different specification for modelling the determinants of inefficiency based on a scaling property,[20] specifically

$$
u_i \sim g(z_i|\delta)u_i^*, \tag{3.23}
$$

where $g(\cdot)$ is a positive function of the exogenous variables (the scaling function) and u_i^* is a positive random variable. With this specification, the distribution of inefficiency is the same for all production units, i.e., governed by u_i^*, while the scale of the inefficiency distribution changes across production units depending on z_i. The scaling property was further explored in Alvarez et al. (2006). Among others, they provided a nice economic interpretation for the scaling property in that u_i^* represents the baseline (in)efficiency of a production unit capturing things like the natural skills of its managers. Meanwhile, the scaling function allows (or prevents) the production unit to exploit these natural skills through other variables, z_i, such as the experience and education of the managers, or the environment in which the production unit operates. Moreover, Alvarez et al. (2006) also devoted their attention to testing the hypothesis of the scaling property.

[20] It is worth mentioning here that although being popularised by Wang and Schmidt (2002), Simar et al. (1994) appear to be the first who analysed the scaling property in detail.

Implementation of Stochastic Frontier Models with Determinants of Inefficiency

Most of the parametric models discussed in this section can easily be implemented using Stata since the estimation routines for the basic stochastic frontier model in Stata also provide options to specify the pre-truncated mean and/or variance of inefficiency as a function of the exogenous variables.

In particular, the Cornwell et al. (1990) estimator can, of course, be implemented using standard panel techniques and linear projections. The model specified in Eq. (3.23) can be estimated using nonlinear least squares without imposing any parametric assumption on the distribution of u_i^* or by the maximum likelihood based on the parametric distribution of the composed error. The maximum likelihood approach can be implemented in Stata by setting up the likelihood using the `sfmodel` command from Kumbhakar et al. (2015) with the option `hscale`$(\cdot)$ and the log likelihood can be maximised using the standard Stata routine `ml max`.

A snippet of Stata codes for implementing stochastic frontier models with determinants of inefficiency is provided in Box 3.8.

Semi-parametric Stochastic Frontier Models

Another generalisation of the stochastic frontier paradigm is to relax parametric assumptions imposed on the functional form of the production frontier and, to some extent, the parametric assumption on the distribution of inefficiency.

The Variety of Models

Banker and Maindiratta (1992) appear to be among the first attempting to estimate stochastic frontier models semi-parametrically. They proposed a framework combining stochastic and deterministic frontier (i.e., data envelopment analysis) approaches and developed techniques for the maximum likelihood estimation with nonparametric characterisation of classes of monotone and concave production frontiers. Other early attempts belong to Fan et al. (1996) and Kneip and Simar (1996), who suggested using nonparametric kernel regression methods in the framework of parametric maximum likelihood estimation. Specifically, Fan

Box 3.8 Illustration for the implementation of stochastic frontier models with determinants of inefficiency

```
*************************************************************************
** Illustration for the implementation of stochastic frontier models ***
**************** with determinants of inefficiency ********************
************************ Partial Stata Codes ***************************
*************************************************************************
/* Note that output, inputs, and exogenous variables are stored in
   global Stata variables $y, $xlist, and $zlist, respectively. Output
   and inputs are in log forms */
/* Implementation of the Kumbhakar et al. (1991) model (using the
   commands from Belotti et al. (2013) ) */
sfcross $y $xlist, distribution(tnormal) emean($zlist)
predict ineff_KGM, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_KGM, bc /* Predict fficiency, i.e., E(exp(-u)|e) */
/* Implementation of the Caudill et al. (1995)  model (using the
   commands from Belotti et al. (2013) ) */
sfcross $y $xlist, distribution(hnormal) usigma($zlist)
predict ineff_CFG, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_CFG, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation of the Wang (2002) model (using the commands from
   Belotti et al. (2013) ) */
sfcross $y $xlist, distribution(tnormal) emean($zlist) usigma($zlist)
predict ineff_W, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_W, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation of the Wang and Schmidt (2002) model (using the
   commands from Kumbhakar et al. (2015) ) */
sfmodel $y, prod dist(t) frontier($xlist) scaling hscale($zlist) tau cu
ml max
sf_predict, jlms(ineff_WH) /* Predict inefficiency, i.e., E(u|e) */
sf_predict, bc(eff_WH) /* Predict efficiency, i.e., E(exp(-u)|e) */
```

et al. (1996) proposed a multi-stage semi-parametric likelihood estimation approach, in which the Nadaraya-Watson nonparametric estimator is employed in the first stage to estimate the average production relationship and a full parametric maximum likelihood estimator is used in the next stage to back out the conditional mean of inefficiency, which is utilised in the last stage to identify the frontier. Kneip and Simar (1996) followed a similar strategy to Fan et al. (1996) but in a panel data setting.

Semi-parametric panel frontiers were also considered in a series of papers by Park et al. (1998, 2003, 2007) wherein firm inefficiency effects are endogenous. They constructed the semi-parametric efficiency bounds and the corresponding semi-parametric efficient estimators for such models under differing assumptions about the form of endogeneity, the serial dependence of the idiosyncratic error, and possible dynamic structures for the panel data model. They used kernel smoothers in these modelling efforts as did Adams et al. (1997, 1999), and Adams and Sickles (2007). Current Stata software for these models is in the development stage in Badunenko et al. (2021), while existing Matlab and R codes for these semi-nonparametric panel frontier models can be found on the website that accompanies Sickles and Zelenyuk (2019) (see the links in footnotes 14 and 15). Model averaging methods utilised in Sickles (2005), Duygun et al. (2017), and Isaksson et al. (2021) also can be found on that website and are currently being used in developing consensus productivity growth estimates for the United Nations Industrial Development Organization (UNIDO).

Another approach to estimate stochastic frontier models semi-parametrically was proposed by Kumbhakar et al. (2007), who suggested employing the local likelihood estimation. The key distinction between this approach and the parametric likelihood approach is that the estimation is localised in the sense that individual contribution to the likelihood is determined by the kernel-based weights instead of the equal weights. Kneip et al. (2015) extended the work of Kumbhakar et al. (2007) by relaxing the parametric assumption on the distribution of inefficiency, while Park et al. (2015) suggested an alternative parameterisation of the local likelihood and outlined a framework for allowing categorical variables in the local likelihood context.

Semi-parametric methods have also been introduced into the stochastic frontier paradigm to deal with specifications of inefficiency. Cornwell et al. (1990) utilised a second-order Taylor series in a time trend to model time-varying inefficiency while Lee and Schmidt (1993) specified the time-varying and cross-sectionally varying inefficiency using a one-factor multiplicative model. Extensions to mixed models and more general factor models were pursued by Ahn et al. (2007, 2013), Kneip et al. (2004, 2012), and Kneip and Sickles (2011). The latter model is programmed in Matlab and R on the software website for Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the links in footnotes 14 and 15) and its coding in Stata is in process in Badunenko et al. (2021).

Finally, the Kneip and Sickles (2011) general cross-sectional and time-varying factor model is available in the R package from Oualid Bada and discussed at length in Bada and Liebl (2014).[21]

Simar et al. (2017) Model

Recently, Simar et al. (2017) suggested using the local least squares method as an alternative for the local likelihood approach to estimate the stochastic frontier models. The local least squares approach is much simpler to compute and easier to implement compared to the local likelihood, and we will focus our discussion here on this approach.

The model in Simar et al. (2017) can be formulated as follows

$$y_i = m(x_i,\ z_i) + v_i - u_i, \quad i = 1, \ldots, n, \tag{3.24}$$

where $m(x_i, z_i)$ is the production frontier, $y_i \in \Re^1_+$ is the output, $x_i \in \Re^p_+$ is a vector of inputs, and $z_i \in \Re^k$ is a vector of k variables that can influence the production process. v_i is statistical noise, which is assumed to have a zero mean, i.e., $E(v_i|x_i, z_i) = 0$, and positive finite variance, i.e., $VAR(v_i|x_i, z_i) \in (0, \infty)$. Meanwhile, u_i is the inefficiency term following a one-sided distribution, with a positive mean, i.e., $E(u_i|x_i, z_i) = \mu_u(x_i, z_i) \in (0, \infty)$ and positive finite variance, i.e., $VAR(u_i|x_i, z_i) \in (0, \infty)$. As in other stochastic frontier models, u_i and v_i are also assumed to be independent, conditionally on (x_i, z_i).

Now, let us define

$$\varepsilon_i^* = v_i - u_i + \mu_u(x_i, z_i), \tag{3.25}$$

and

$$r_1(x_i, z_i) = m(x_i, z_i) - \mu_u(x_i, z_i). \tag{3.26}$$

We can rewrite (3.24) as

$$y_i = r_1(x_i, z_i) + \varepsilon_i^*. \tag{3.27}$$

Since $E\left(\varepsilon_i^*|x_i, z_i\right) = 0$, we can use standard nonparametric methods (e.g., local polynomial least squares) to estimate $r_1(x_i, z_i)$. In order to estimate

[21] Software instructions and downloadable codes are accessible at https://www.jstatsoft.org/article/view/v059i06.

the individual inefficiency, we also need to make a parametric assumption on the distribution of inefficiency, e.g.,

$$u_i|x_i, z_i \sim \mathcal{N}^+\Big(0, \sigma_u^2(x_i, z_i)\Big). \tag{3.28}$$

With the distributional assumption, the conditional mean of inefficiency can be estimated using the following relationships

$$\sigma_u^3(x_i, z_i) = \sqrt{\frac{\pi}{2}}\left(\frac{\pi}{\pi - 4}\right) r_3(x_i, z_i), \tag{3.29}$$

and

$$\mu_u(x_i, z_i) = \sqrt{\frac{2}{\pi}}\sigma_u(x_i, z_i), \tag{3.30}$$

where $r_3(x_i, z_i) = E\Big(\big(\varepsilon_i^*\big)^3 | x_i, z_i\Big)$ is the third moment of the composed error. Specifically, the residuals from the nonparametric estimation of the model in Eq. (3.27), $\hat{\varepsilon}_i^*$, can be utilised to obtain the nonparametric estimates of the third moment of the composed error, $\hat{r}_3(x_i, z_i)$. The estimates of technical inefficiency then can be obtained by plugging the $\hat{r}_3(x_i, z_i)$ into Eqs. (3.29) and (3.30).[22]

Implementation of Simar et al. (2017) Model

Estimation of the Simar et al. (2017) model can be implemented using the standard Stata routines with a bit of additional programming. The key command is `npregress` which helps to perform the local least-square estimation in the Stata environment. As an illustration, we provide here, in Box 3.9, a part of a Stata do file that implements the procedure discussed in the previous subsection to estimate the Simar et al. (2017) model.

Similarly, one can implement the estimation of the Simar et al. (2017) model in R with the local least squares estimation being carried out by the

[22] The distributional assumptions on u_i and v_i allow obtaining a generalised version of JLMS-type estimates, although more interesting in the semi/non-parametric context are the estimates of $E(u_i|x_i = x, z_i = z)$, which can be done for any values of interest for (x, z). The elasticities of $E(u_i|x_i = x, z_i = z)$ can also be obtained, which can be done without any parametric assumptions on distributions, just by assuming that u_i comes from a one-parameter scale family (see Sect. 4 in Simar et al. 2017 for more details).

Box 3.9 Illustration for the implementation of the Simar et al. (2017) model

```
**********************************************************************
* Illustration for the implementation of the Simar et al. (2017) model**
***************************Partial Stata Codes************************
**********************************************************************
/* Estimate the model in (27) using local linear estimator and store
   predicted value in variable r1hat. Note that output, inputs, and
   exogenous variables are stored in global Stata variables $y, $xlist,
   and $zlist, respectively. Output and inputs are in log forms */
/* Note that the default options in the npregress command is to use
   Epanechnikov kernel and select bandwidth by cross-validation, i.e.,
   by minimizing the integrated mean squared error of the prediction. */
npregress kernel $y $xlist $zlist, estimator(linear) predict(r1hat)
   noderivatives
/* Obtain the residual and the residual cubed from  estimation of the
   model in equation (27)*/
gen ehat = $y - r1hat
gen ehat3 = ehat^3
/* Estimate the third moment of the composed error using local linear
   estimator and store predicted value in variable r3hat */
npregress kernel ehat3 $xlist $zlist, estimator(linear) predict(r3hat)
   noderivatives
/* Calculate sigma u hat cubed using equation (29) */
gen sigmauhat3 = sqrt(_pi/2)*(_pi/(_pi-4))*r3hat
/* Calculate sigma u hat. Note that following Simar et al. (2017), we
   set negative values of sigma u hat equal zero */
gen sigmauhat = max(sigmauhat3^(1/3),0)
/* Calculate estimated values of inefficiency using equation (30)*/
gen muhat = sqrt(2/_pi)*sigmauhat
```

np package with a bit of additional programming similar to the one we presented here (and as was done by Parmeter and Zelenyuk [2019]). The implementation of the model in Matlab requires more effort since one needs to write his/her own codes for the local least square estimation (as was done by Simar et al. [2017]). Preparation of user-friendly packages in R and Matlab is currently in progress.

Empirical Illustration

In this section, we provide a small empirical illustration of the models discussed in the previous sections, including the basic stochastic frontier model, the stochastic panel data models, and the semi-parametric stochastic frontier model.[23,24] For this purpose, we use the data set about rice producers in the Philippines, which was also utilised for the similar purpose and popularised in the literature by Coelli et al. (2005).[25,26]

Specifically, the data set includes the information about 43 rice producers in the Tarlac region of the Philippines in a period of 8 years from 1990 to 1997. We extract from the data set the information on one output and three inputs including the area planted, labour used, and fertiliser used. The output is measured in tonnes of freshly threshed rice, while the inputs are measured in hectares, man-days of family and hired labour, and kilograms of active ingredients, respectively (see more details about the description of the data in Coelli et al., 2005).

For this empirical illustration, we deliberately apply all the models to the data and focus our discussion on the estimated inefficiency to reflect the differences in results across the models. Moreover, for all the models that require a functional form for the production relationship, we assume a linear in log production function, i.e., the Cobb-Douglas production function.[27] The Stata codes for implementing this analysis are provided in the Appendix.

[23] For the results to some extent to be comparable, we deliberately do not include in this empirical illustration the stochastic frontier models with determinants of inefficiency.

[24] Also, due to the computational difficulty in optimising the likelihood function, the result from Kumbhakar (1990) is not available for the dataset used in this empirical illustration.

[25] Downloaded from http://www.uq.edu.au/economics/cepa/crob2005/software/CROB2005.zip.

[26] For an illustration with this data with various DEA models see, e.g., Simar and Zelenyuk (2020).

[27] To estimate the cross-sectional models, e.g., Aigner et al. (1977) and Simar et al. (2017) models, we pool the data across years.

Table 3.1 Summary statistics of the estimated inefficiency

Models	*Mean*	*Std. dev.*	*Min*	*Q1*	*Median*	*Q3*	*Max*
Aigner et al. (1977)	0.36	0.25	0.04	0.18	0.29	0.47	2.00
Schmidt and Sickles (1984) (fixed effects)	0.34	0.20	0.00	0.18	0.30	0.50	0.98
Schmidt and Sickles (1984) (random effects)	0.23	0.12	0.00	0.13	0.21	0.33	0.60
Pitt and Lee (1981)	0.21	0.14	0.05	0.10	0.16	0.31	0.70
Cornwell et al. (1990)	0.44	0.27	0.00	0.27	0.41	0.57	2.09
Battese and Coelli (1992)	0.20	0.15	0.04	0.08	0.14	0.28	0.92
Greene (2005a, 2005b) (random effects)	0.33	0.24	0.03	0.17	0.27	0.43	1.89
Greene (2005a, 2005b) (fixed effects)	0.35	0.23	0.02	0.18	0.30	0.47	1.87
Kumbhakar et al. (2014) (Total)	0.45	0.25	0.10	0.27	0.38	0.55	2.01
Kumbhakar et al. (2014) (Persistent)	0.15	0.10	0.02	0.06	0.13	0.24	0.49
Kumbhakar et al. (2014) (Transitory)	0.29	0.20	0.03	0.16	0.24	0.37	1.67
Simar et al. (2017)	0.25	0.14	0.00	0.17	0.30	0.35	0.45

The summary statistics of the estimated inefficiency are provided in Table 3.1, and their histograms are shown in Fig. 3.1.[28] Meanwhile, the variations of the estimated inefficiency across the years are shown in Fig. 3.2.

At first glance, we can see that the means of estimated inefficiency vary significantly across the models, ranging from 0.20 (the Battese and Coelli [1992] model) to 0.45 (the Kumbhakar et al. [2014] model).[29] This is understandable since each model depends on different sets of assumptions. Moreover, it is important for practitioners to be aware of these differences and carefully justify the assumptions of the model of their choice before proceeding with their analysis. For example, with this data

[28] The estimated distribution of estimated inefficiency from the Simar et al. (2017) model is showing some mass at zero (i.e., the phenomenon referred to as "wrong skewness" in stochastic frontier analysis) because 79 out 344 observations have $\widehat{\sigma}_u^3(x_i, z_i) < 0$ and their inefficiency is set to equal to 0.

[29] It is important to clarify here that for all the models, the means we refer to are averages of the estimates of individual inefficiencies.

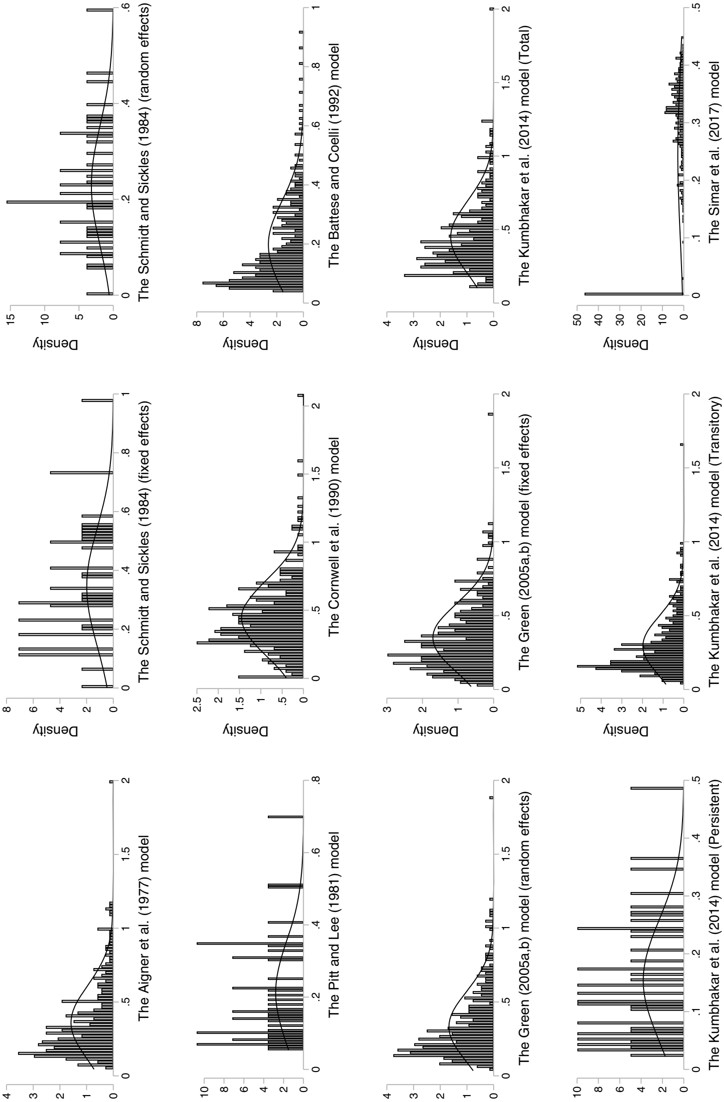

Fig. 3.1 Histograms of the estimated inefficiency

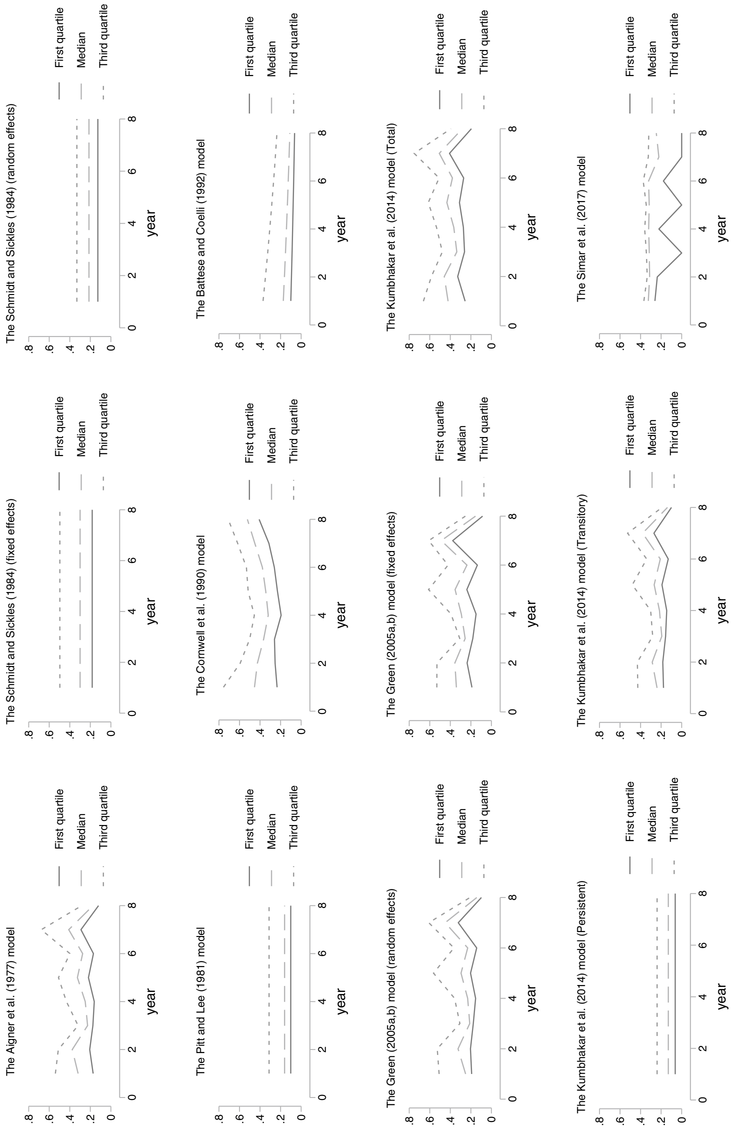

Fig. 3.2 The estimated inefficiency across years

set, the difference in estimated inefficiency between the fixed effects and random effects frameworks is significant when all the unobserved individual heterogeneity is viewed as inefficiency (e.g., in the Schmidt and Sickles [1984] model), but the difference is minimal when inefficiency is distinguished from the unobserved individual heterogeneity (e.g., in the Greene [2005a, 2005b] model).

Furthermore, recall that the temporal pattern of inefficiency is also specified differently in different models. As illustrated in Fig. 3.2, the estimated inefficiency is constant over time in the Schmidt and Sickles (1984) and Pitt and Lee (1981) models, but follows a quadratic trend in the Cornwell et al. (1990) model and has a linear trend in Battese and Coelli (1992). Meanwhile, other models, such as Greene (2005a, 2005b) and Kumbhakar et al. (2014), do not impose any temporal patterns on the time-varying component of inefficiency.

Concluding Remarks

This chapter discussed a variety of stochastic frontier models to estimate the technical efficiency of production units. Our chapter also documented the estimation routines used to implement these methods for practitioners, especially those who are willing to use Stata, but also with tips on where to find analogous programs for R and Matlab users.

Although many recent developments in the field were covered in this chapter, it was still a relatively brief introduction to the stochastic frontier paradigm with some other generalisations remaining untouched, such as the Bayesian stochastic frontier,[30] stochastic metafrontier,[31] spillovers and spatial frontiers,[32] and endogeneity.[33] We refer interested readers to more extensive resources (e.g., Kumbhakar et al., 2021a, 2021b; Sickles & Zelenyuk, 2019) for more detailed discussions of these and other topics.

[30] For example, see Van den Broeck et al. (1994), Griffin and Steel (2004, 2007), and Liu et al. (2017).

[31] For example, see Battese et al. (2004), O'Donnell et al. (2008), and Huang et al. (2014).

[32] For example, see Glass et al. (2016), Orea and Álvarez (2019).

[33] For example, see Amsler et al. (2016), Kutlu (2010), Karakaplan and Kutlu (2015, 2017), and Karakaplan (2017).

Finally, many other important developments in the field are still in progress, and thus, we encourage readers to check for updates as well as contribute themselves to such developments and discoveries.

Acknowledgements We thank the Editor and two anonymous referees for fruitful comments. Bao Hoang Nguyen and Valentin Zelenyuk acknowledge the support from the University of Queensland and the financial support from the Australian Research Council (FT170100401). We also thank Hong Ngoc Nguyen, Zhichao Wang, Evelyn Smart, Anne-Claire Bouton, and Travis Alan Smith for their valuable feedback. These individuals and organisations are not responsible for the views expressed in this paper.

Appendix

Box A.1: Stata codes for the empirical illustration

```
set more off
clear all
log using "SFABookChapter_Empirical", text replace
// Load data
import delimited "rice.txt", delimiter(space) varnames(1) ///
        encoding(ISO-8859-2)
// Generate variables
foreach X of varlist prod area labor npk {
        generate l`X' = ln(`X')
        }
global y lprod
global xlist llabor larea lnpk  /*Cobb-Douglas function*/
global id fmercode
global t yeardum
xtset $id $t
************************************************************************
/* the Aigner et al. (1977) model */
************************************************************************
sfcross $y $xlist, distribution(hnormal)
estimates store ALS
predict ineff_ALS, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_ALS "The Aigner et al. (1977) model"
************************************************************************
/* the  Schmidt and Sickles (1984) model        */
****************************Partial Stata Codes*************************
/* the fixed effect framework */
sfpanel $y $xlist, model(fe)
estimates store SS_fe
predict ineff_SS_fe, u /* Predict inefficiency*/
label variable ineff_SS_fe ///
        "The Schmidt and Sickles (1984)(fixed effects)"
/* the random effect framework */
sfpanel $y $xlist, model(regls)
estimates store SS_re
predict ineff_SS_re, u /* Predict inefficiency*/
label variable ineff_SS_re ///
        "The Schmidt and Sickles (1984) (random effects)"
```

```
************************************************************************
/* the Pitt and Lee (1981) model */
************************************************************************
sfpanel $y $xlist, model(pl81)
estimates store PL
predict ineff_PL, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_PL "The Pitt and Lee (1981) model"
************************************************************************
/* the Cornwell et al. (1990) model */
************************************************************************
sfpanel $y $xlist, model(fecss)
estimates store CSS
predict ineff_CSS, u /* Predict inefficiency*/
label variable ineff_CSS "The Cornwell et al. (1990) model"
************************************************************************
/* the Battese and Coelli (1992) model */
************************************************************************
sfpanel $y $xlist, model(bc92)
estimates store BC
predict ineff_BC, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_BC "The Battese and Coelli (1992) model"
************************************************************************
/* the Green (2005a,b) models */
************************************************************************
/* the random effect framework */
sfpanel $y $xlist, distribution(hnormal) model(tre)
estimates store G_tre
predict ineff_G_tre, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_G_tre "The Green (2005a,b) model (random effects)"
/* the fixed effect framework */
/* Implementation using the marginal maximum likelihood estimator */
sftfe $y $xlist,  distribution(hnormal) estimator(within)
estimates store G_mmle
predict ineff_G_mmle, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_G_mmle "The Green (2005a,b) model (fixed effects)"
************************************************************************
/* the Kumbhakar et al. (2014) models */
************************************************************************
xtreg $y $xlist, re
estimates store KLH
predict alp, u /* Obtain estimates of alpha */
predict esl, e /* Obtain estimates of the composed error */
/* Estimate equation (19) using the basic stochastic frontier model to
   obtain persistent (in)efficiency */
```

```
gen constant = 1 /* Generate constant to use in sfcross */
sfcross alp constant, distribution(hnormal) noconstant
predict ineff_KHL_pers, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_KHL_pers ///
        "The Kumbhakar et al. (2014) model (Persistent)"
/* Estimate equation (19) using the basic stochastic frontier model to
   obtain transitory (in)efficiency */
sfcross esl constant, distribution(hnormal) noconstant
predict ineff_KHL_trans, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_KHL_trans ///
        "The Kumbhakar et al. (2014) model (Transitory)"
/*Calculate the total inefficiency*/
gen ineff_KLH = ineff_KHL_pers + ineff_KHL_trans
label variable ineff_KLH "The Kumbhakar et al. (2014) model (Total)"
***********************************************************************
/* the Simar et al. (2017) */
***********************************************************************
/* Estimate model in (27) using local linear estimator and store
   predicted value in variable r1hat. Note that output, inputs, and
   exogenous variables are stored in global Stata variables $y, $xlist,
   and $zlist, respectively */
/* Note that the default options in the npregress command is to use
   Epanechnikov kernel and select bandwidth by cross-validation, i.e.,
   by minimizing the integrated mean squared error of the prediction. */
npregress kernel $y $xlist, estimator(linear) predict(r1hat) ///
        noderivatives
/* Obtain the residual and the residual cubed from the estimation of the
    model in equation (27)*/
gen ehat = $y - r1hat
gen ehat3 = ehat^3
/* Estimate the third moment of the composed error using local linear
   estimator and store predicted value in variable r3hat */
npregress kernel ehat3 $xlist, estimator(linear) predict(r3hat) ///
        noderivatives
/* Calculate sigma u hat cubed using equation (29) */
gen sigmauhat3 = sqrt(_pi/2)*(_pi/(_pi-4))*r3hat
/* Calculate sigma u hat. Note that following Simar et al. (2017), we
   set negative values of sigma u hat equal zero */
gen sigmauhat = max(sigmauhat3^(1/3),0)
/* Calculate estimated values of inefficiency using equation (30)*/
gen ineff_SKVZ = sqrt(2/_pi)*sigmauhat
label variable ineff_SKVZ "The Simar et al. (2017) model"
```

```
*********************************************************************
/* Summarising and exporting the results */
*********************************************************************
/* The estimated coefficients of the frontiers */
esttab ALS SS_fe SS_re PL CSS BC G_tre G_mmle KLH
esttab ALS SS_fe SS_re PL CSS BC G_tre G_mmle KLH  ///
        using coefficients.csv, replace
/* The estimated inefficiency */
global myvars ineff_ALS ineff_SS_fe ineff_SS_re ineff_PL ineff_CSS ///
        ineff_BC ineff_G_tre ineff_G_mmle ineff_KLH ineff_KHL_pers ///
        ineff_KHL_trans ineff_SKVZ
estpost summarize $myvars, detail
esttab using inefficiency.csv, ///
        cells("count mean sd min p25 p50 p75 max") replace
*********************************************************************
/* Histograms of estimated inefficiency */
*********************************************************************
foreach X of varlist $myvars {
        histogram `X', bin(100) normal `kden'
        graph save `X'.gph, replace
        }
graph combine ineff_ALS.gph ineff_SS_fe.gph ineff_SS_re.gph ///
        ineff_PL.gph ineff_CSS.gph ineff_BC.gph ineff_G_tre.gph ///
        ineff_G_mmle.gph ineff_KLH.gph ineff_KHL_pers.gph ///
        ineff_KHL_trans.gph ineff_SKVZ.gph, col(3) scale(1)
graph export histogramineff.png, replace
*********************************************************************
/* Plot estimated inefficiency across years */
*********************************************************************
sort yeardum
label variable yeardum "year"
foreach X of varlist $myvars {
        by yeardum, sort: egen `X'_Q1 = pctile(`X'), p(25)
        by yeardum, sort: egen `X'_Q2 = pctile(`X'), p(50)
        by yeardum, sort: egen `X'_Q3 = pctile(`X'), p(75)
        label variable `X'_Q1 "First quartile"
        label variable `X'_Q3 "Third quartile"
        label variable `X'_Q2 "Median"
        local labeltext : variable label `X'
        graph two line `X'_Q1 `X'_Q2 `X'_Q3 yeardum, ///
                title(`labeltext', size(small))
        graph save `X'_trend.gph, replace
        }
graph combine ineff_ALS_trend.gph ineff_SS_fe_trend.gph ///
```

```
        ineff_SS_re_trend.gph ineff_PL_trend.gph ///
        ineff_CSS_trend.gph ineff_BC_trend.gph ///
        ineff_G_tre_trend.gph ineff_G_mmle_trend.gph ///
        ineff_KLH_trend.gph ineff_KHL_pers_trend.gph ///
        ineff_KHL_trans_trend.gph ineff_SKVZ_trend.gph, ///
        col(3) scale(1) xcommon ycommon
graph export allineff_trend.png, replace
log close
```

References

Adams, R. M., Berger, A. N., & Sickles, R. C. (1997). Computation and inference in semiparametric efficient estimation. In H. Amman, B. Rustem, & A. Whinston (Eds.), *Computational Approaches to Economic Problems* (pp. 57–70). Springer.

Adams, R. M., Berger, A. N., & Sickles, R. C. (1999). Semiparametric approaches to stochastic panel frontiers with applications in the banking industry. *Journal of Business & Economic Statistics, 17*(3), 349–358.

Adams, R. M., & Sickles, R. C. (2007). Semiparametric efficient distribution free estimation of panel models. *Communications in Statistics-Theory and Methods, 36*(13), 2425–2442.

Ahn, S. C., Lee, Y. H., & Schmidt, P. (2007). Stochastic frontier models with multiple time-varying individual effects. *Journal of Productivity Analysis, 27*(1), 1–12.

Ahn, S. C., Lee, Y. H., & Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics, 174*(1), 1–14.

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 64*(6), 1263–1297.

Almanidis, P., Qian, J., & Sickles, R. C. (2014). Stochastic frontier models with bounded inefficiency. In R. C. Sickles & W. C. Horrace (Eds.), *Festschrift in honor of Peter Schmidt: Econometric methods and applications* (pp. 47–81). Springer.

Almanidis, P., & Sickles, R. C. (2012). The skewness problem in stochastic frontier models: Fact or fiction? In I. V. Keilegom & P. W. Wilson (Eds.), *Exploring research frontiers in contemporary statistics and econometrics: A Festschrift in honor of Léopold Simar* (pp. 201–227). Springer.

Alvarez, A., Amsler, C., Orea, L., & Schmidt, P. (2006). Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis, 25*(3), 201–212.

Álvarez, I. C., Barbero, J., & Zofío, J. L. (2017). A panel data toolbox for matlab. *Journal of Statistical Software, 76*(6), 1–27.
Amsler, C., Prokhorov, A., & Schmidt, P. (2016). Endogeneity in stochastic frontier models. *Journal of Econometrics, 190*(2), 280–288.
Bada, O., & Liebl, D. (2014). phtt: Panel data analysis with heterogeneous time trends in R. *Journal of Statistical Software, 59*(6), 1–33.
Badunenko, O., Sickles, R. C., & Zelenyuk, V. (2021). *Productivity and efficiency analysis in Stata*, monograph in progress. Rice University.
Banker, R. D., & Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. *Journal of Productivity Analysis, 3*(4), 401–415.
Battese, G. E., & Broca, S. S. (1997). Functional forms of stochastic frontier production functions and models for technical inefficiency effects: A comparative study for wheat farmers in Pakistan. *Journal of Productivity Analysis, 8*(4), 395–414.
Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics, 38*(3), 387–399.
Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis, 3*(1–2), 153–169.
Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics, 20*(2), 325–332.
Battese, G. E., Rao, D. S. P., & O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis, 21*(1), 91–103.
Belotti, F., Daidone, S., Ilardi, G., & Atella, V. (2013). Stochastic frontier analysis using Stata. *The Stata Journal, 13*(4), 719–758.
Belotti, F., & Ilardi, G. (2018). Consistent inference in fixed-effects stochastic frontier models. *Journal of Econometrics, 202*(2), 161–177.
Bogetoft, P., & Otto, L. (2019). *Benchmarking: Benchmark and frontier analysis using DEA and SFA* (R package version 0.28).
Caudill, S. B., Ford, J. M., & Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business & Economic Statistics, 13*(1), 105–111.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.
Chen, Y.-Y., Schmidt, P., & Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics, 181*(2), 65–76.

Coelli, T. J., & Henningsen, A. (2020). *Package 'frontier'* (R package version 1.1-8).

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer.

Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis, 42*(2), 123–136.

Comans, T., Nguyen, K.-H., Zelenyuk, V., Gray, L., Tran, A., Nguyen, B. H., Wang, Z., & Moretto, N. (2020). *The cost of residential aged care* (Technical Report Research Paper 9). The Royal Commission into Aged Care Quality and Safety.

Cornwell, C., Schmidt, P., & Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics, 46*(1–2), 185–200.

Daraio, C., Kerstens, K. H., Nepomuceno, T. C. C., & Sickles, R. C. (2019). Productivity and efficiency analysis software: An exploratory bibliographical survey of the options. *Journal of Economic Surveys, 33*(1), 85–100.

Daraio, C., Kerstens, K. H., Nepomuceno, T. C. C., & Sickles, R. C. (2020). Empirical surveys of frontier applications: A meta-review. *International Transactions in Operational Research, 27*(2), 709–738.

Duygun, M., Hao, J., Isaksson, A., & Sickles, R. C. (2017). World productivity growth: A model averaging approach. *Pacific Economic Review, 22*(4), 587–619.

Fan, Y., Li, Q., & Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business & Economic Statistics, 14*(4), 460–468.

Färe, R., He, X., Li, S., & Zelenyuk, V. (2019). A unifying framework for Farrell profit efficiency measurement. *Operations Research, 67*(1), 183–197.

Farrell, M. J. (1957). The measurement of productive efficiency. *Series A (General): Journal of the Royal Statistical Society, 120*(3), 253–290.

Fenochietto, R., & Pessino, C. (2013). *Understanding countries' tax effort* (IMF Working Paper).

Ferrier, G. D., & Lovell, C. K. (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of Econometrics, 46*(1–2), 229–245.

Glass, A., Kenjegalieva, K., & Sickles, R. C. (2016). A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers. *Journal of Econometrics, 190*(2), 289–300.

Greene, W. H. (1980a). Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics, 13*(1), 27–56.

Greene, W. H. (1980b). On the estimation of a flexible frontier production model. *Journal of Econometrics, 13*(1), 101–115.

Greene, W. H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics, 46*(1), 141–163.
Greene, W. H. (2004). Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics, 13*(10), 959–980.
Greene, W. H. (2005a). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis, 23*(1), 7–32.
Greene, W. H. (2005b). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics, 126*(2), 269–303.
Griffin, J. E., & Steel, M. F. (2004). Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics, 123*(1), 121–152.
Griffin, J. E., & Steel, M. F. (2007). Bayesian stochastic frontier analysis using winbugs. *Journal of Productivity Analysis, 27*(3), 163–176.
Huang, C. J., Huang, T.-H., & Liu, N.-H. (2014). A new approach to estimating the metafrontier production function based on a stochastic frontier framework. *Journal of Productivity Analysis, 42*(3), 241–254.
Isaksson, A., Shang, C., & Sickles, R. C. (2021). Nonstructural analysis of productivity growth for the industrialized countries: A jackknife model averaging approach. *Econometric Reviews, 40*(4), 321–358.
Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics, 19*(2–3), 233–238.
Karakaplan, M. U. (2017). Fitting endogenous stochastic frontier models in Stata. *The Stata Journal, 17*(1), 39–55.
Karakaplan, M. U., & Kutlu, L. (2015). *Handling endogeneity in stochastic frontier analysis*. Available at SSRN 2607276.
Karakaplan, M. U., & Kutlu, L. (2017). Endogeneity in panel stochastic frontier models: An application to the Japanese cotton spinning industry. *Applied Economics, 49*(59), 5935–5939.
Kneip, A., & Sickles, R. C. (2011). Panel data, factor models, and the Solow residual. In I. Van Keilegom & P. W. Wilson (Eds.), *Exploring research frontiers in contemporary statistics and econometrics: A Festschrift for Léopold Simar* (pp. 83–114). Physica.
Kneip, A., Sickles, R. C., & Song, W. (2004). Functional data analysis and mixed effect models. In J. Antoch (Ed.), *COMPSTAT 2004—Proceedings in Computational Statistics* (pp. 315–326). Physica.
Kneip, A., Sickles, R. C., & Song, W. (2012). A new panel data treatment for heterogeneity in time trends. *Econometric Theory, 28*(3), 590–628.
Kneip, A., & Simar, L. (1996). A general framework for frontier estimation with panel data. *Journal of Productivity Analysis, 7*(2), 187–212.

Kneip, A., Simar, L., & Van Keilegom, I. (2015). Frontier estimation in the presence of measurement error with unknown variance. *Journal of Econometrics, 184*(2), 379–393.
Kumbhakar, S. C. (1987). The specification of technical and allocative inefficiency in stochastic production and profit frontiers. *Journal of Econometrics, 34*(3), 335–348.
Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics, 46*(1–2), 201–211.
Kumbhakar, S. C., Ghosh, S., & McGuckin, J. T. (1991). A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business & Economic Statistics, 9*(3), 279–286.
Kumbhakar, S. C., Lien, G., & Hardaker, J. B. (2014). Technical efficiency in competing panel data models: A study of Norwegian grain farming. *Journal of Productivity Analysis, 41*(2), 321–337.
Kumbhakar, S. C., Park, B. U., Simar, L., & Tsionas, E. G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics, 137*(1), 1–27.
Kumbhakar, S. C., Parmeter, C. F., & Zelenyuk, V. (2021a). Stochastic frontier analysis: Foundations and advances I. In S. C. Ray, R. Chambers, & S. C. Kumbhakar (Eds.), *Handbook of production economics* (pp. 1–40). Springer Singapore.
Kumbhakar, S. C., Parmeter, C. F., & Zelenyuk, V. (2021b). Stochastic frontier analysis: Foundations and advances II. In S. C. Ray, R. Chambers, & S. C. Kumbhakar (Eds.), *Handbook of production economics* (pp. 1–38). Springer Singapore.
Kumbhakar, S. C., & Tsionas, E. G. (2005). Measuring technical and allocative inefficiency in the translog cost system: A Bayesian approach. *Journal of Econometrics, 126*(2), 355–384.
Kumbhakar, S. C., & Tsionas, E. G. (2008). Estimation of input-oriented technical efficiency using a nonhomogeneous stochastic production frontier model. *Agricultural Economics, 38*(1), 99–108.
Kumbhakar, S. C., Wang, H., & Horncastle, A. P. (2015). *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.
Kutlu, L. (2010). Battese-Coelli estimator with endogenous regressors. *Economics Letters, 109*(2), 79–81.
Lee, Y. H., & Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical efficiency. In H. O. Fried, S. S. Schmidt, & C. K. Lovell (Eds.), *The measurement of productive efficiency: Techniques and applications* (pp. 237–255). Oxford University Press.
Liu, J., Sickles, R. C., & Tsionas, E. G. (2017). Bayesian treatments for panel data stochastic frontier models with time varying heterogeneity. *Econometrics, 5*(3), 1–21.

Malikov, E., Kumbhakar, S. C., & Tsionas, M. G. (2016). A cost system approach to the stochastic directional technology distance function with undesirable outputs: The case of US banks in 2001–2010. *Journal of Applied Econometrics, 31*(7), 1407–1429.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review, 18*(2), 435–444.

Mutter, R. L., Greene, W. H., Spector, W., Rosko, M. D., & Mukamel, D. B. (2013). Investigating the impact of endogeneity on inefficiency estimates in the application of stochastic frontier analysis to nursing homes. *Journal of Productivity Analysis, 39*(2), 101–110.

Nguyen, B. H., & Zelenyuk, V. (2021), Aggregation of outputs and inputs for DEA analysis of hospital efficiency: Economics, operations research and data science perspectives. In J. Zhu & V. Charles (Eds.), *Data-enabled analytics: DEA for big data* (pp. 123–158). Springer.

O'Donnell, C. J., Rao, D. S. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics, 34*(2), 231–255.

Orea, L., & Álvarez, I. C. (2019). A new stochastic frontier model with cross-sectional effects in both noise and inefficiency terms. *Journal of Econometrics, 213*(2), 556–577.

Park, B. U., Sickles, R. C., & Simar, L. (1998). Stochastic panel frontiers: A semiparametric approach. *Journal of Econometrics, 84*(2), 273–301.

Park, B. U., Sickles, R. C., & Simar, L. (2003). Semiparametric-efficient estimation of AR(1) panel data models. *Journal of Econometrics, 117*(2), 279–309.

Park, B. U., Sickles, R. C., & Simar, L. (2007). Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics, 136*(1), 281–301.

Park, B. U., Simar, L., & Zelenyuk, V. (2015). Categorical data in local maximum likelihood: Theory and applications to productivity analysis. *Journal of Productivity Analysis, 43*(2), 199–214.

Parmeter, C. F., & Zelenyuk, V. (2019). Combining the virtues of stochastic frontier and data envelopment analysis. *Operations Research, 67*(6), 1628–1658.

Pitt, M. M., & Lee, L.-F. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics, 9*(1), 43–64.

Rosko, M. D. (2001). Cost efficiency of US hospitals: A stochastic frontier approach. *Health Economics, 10*(6), 539–551.

Schmidt, P., & Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics, 2*(4), 367–374.

Sickles, R. C. (2005). Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. *Journal of Econometrics, 126*(2), 305–334.

Sickles, R. C., Song, W., & Zelenyuk, V. (2020). Econometric analysis of productivity: Theory and implementation in R. Handbook of Statistics. In H. D. Vinod & C. R. Rao (Eds.), *Financial, macro and micro econometrics using R* (Vol. 42, pp. 267–297). Elsevier.

Sickles, R. C., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency*. Cambridge University Press.

Simar, L., Lovell, C. K., & van den Eeckaut, P. (1994). *Stochastic frontiers incorporating exogenous influences on efficiency* (STAT Discussion Papers).

Simar, L., Van Keilegom, I., & Zelenyuk, V. (2017). Nonparametric least squares methods for stochastic frontier models. *Journal of Productivity Analysis, 47*(3), 189–204.

Simar, L., & Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis, 36*(1), 1–20.

Simar, L., & Zelenyuk, V. (2020). Improving finite sample approximation by central limit theorems for estimates from data envelopment analysis. *European Journal of Operational Research, 284*(3), 1002–1015.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics, 13*(1), 57–66.

Van den Broeck, J., Koop, G., Osiewalski, J., & Steel, M. F. (1994). Stochastic frontier models: A Bayesian perspective. *Journal of Econometrics, 61*(2), 273–303.

Wang, H. J. (2002). Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis, 18*(3), 241–253.

Wang, H. J., & Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis, 157*, 286–296.

Zelenyuk, V. (2020). Aggregation of inputs and outputs prior to data envelopment analysis under big data. *European Journal of Operational Research, 282*(1), 172–187.

Zuckerman, S., Hadley, J., & Lezzoni, L. (1994). Measuring hospital efficiency with frontier cost functions. *Journal of Health Economics, 13*(3), 255–280.

CHAPTER 4

Efficiency and Productivity Analysis from a System Perspective: Historical Overview

Antonio Peyrache and Maria C. A. Silva

Prologo

In this chapter, we focus on a particular branch of efficiency and productivity analysis that mostly relates to Network Data Envelopment Analysis

We would like to thank Knox Lovell and Prasada Rao for reading a version of this paper and providing comments and insights.

A. Peyrache (✉)
Centre for Efficiency and Productivity Analysis (CEPA), School of Economics, University of Queensland, St Lucia, QLD, Australia
e-mail: a.peyrache@uq.edu.au

M. C. A. Silva
CEGE—Católica Porto Business School, Porto, Portugal
e-mail: csilva@ucp.pt

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_4

(NDEA) models in their connection to what has been called the centralized allocation model or industry efficiency model. Both of these models may be thought as being part of an analytical approach that looks at productivity and efficiency analysis from a system perspective rather than the more traditional granular perspective of plant or firm efficiency analysis. From this point of view, the models can be better connected with issues of regulation of markets that present strong externalities or distortions, or issues of efficient allocation of limited resources in government centrally planned operations. The reason why we focus on NDEA models in particular is due to their astonishing growth in the last 5 to 10 years. A Google Scholar search dated 24/02/2021 with either "Network DEA" or "Network Data Envelopment Analysis" in the title returns 887 research papers. By limiting the same search to before year 1999, one obtains zero papers. Between year 2000 and 2005, 9 papers were published. Between year 2006 and 2010, 87 papers were published. Between 2011 and 2015, 252 papers were published. After 2015 until today, 572 papers have been published. This is an astonishingly exponential growth of what was a tiny little detail in productivity analysis. This search does not include papers that include "Network DEA" or "Network Data Envelopment Analysis" outside of the title. If we remove the requirement for these two sentences to appear in the title, 7,520 papers appear from the search, with a similar temporal distribution: 54 papers before 1999, 92 papers between 2000 and 2005, 419 papers between 2006 and 2010, 1,770 papers between 2011 and 2015, and 5,060 papers between 2016 and 2021. This is a huge amount of papers for such a specialized topic and, to the best of our knowledge, no other sub-field in efficiency and productivity analysis has undergone such miraculous growth. One is therefore left with a feeling of backwardness, as if the modern researcher in productivity and efficiency analysis is missing the biggest leap forward in our knowledge of the field. This motivated us to make a very selective review of this large body of literature. During this process, we stumbled across the contributions of Kantorovich (1939, 1965), Koopmans (1951) and Johansen (1972) and we formed the view that this field of study is far from being a specialized field within efficiency and productivity analysis, but it is rather the best effort to make a connection with economic policy issues associated with central planning and the regulation of markets. Since it is tedious, boring, and almost impossible to review all of these papers, we decided to focus on papers that received the highest number of citations, with a special focus on papers published after 2015. Having a bit of a bigger focus on what happened after 2015 would help in mitigating the distortions that could arise by the citation game. Although this is not necessarily the best way of reviewing the literature and there could be very good papers that received a small number of citations, we nevertheless decided to proceed this way. From the above search, we selected a bit more than 150 papers that we reviewed in order to gain an understanding of what is happening in the field. This chapter is an attempt at explaining in a succinct way our view of this growing body of literature (and we cite, from those 150, only papers that we think are relevant to our discussion, without having the ambition of providing an exhaustive literature review). During our search, we developed our independent modeling strategy to try to reconcile these papers. The outcome of this modeling strategy is contained in Peyrache and Silva (2019).

The origins of system models in efficiency and productivity analysis can be traced back to Kantorovich (1939). In essence, a system is a set of interacting or interdependent group of items forming a unified whole. The system has properties that its parts do not necessarily possess. As Senge (1990) mentions in his system thinking approach: a plane can fly while none of its parts can. Under production economics, systems can be considered groups of firms acting in an industry, or production processes acting within a firm.

Farrell (1957) is often cited as the father of modern efficiency and productivity analysis either through parametric or nonparametric techniques. In his seminal paper, he mentions the measurement of industry efficiency in the following words:

> There is, however, a very satisfactory way of getting round this problem: that is, by comparing an industry's performance with the efficient production function derived from its constituent firms. The 'technical efficiency' of an industry measured in this way, will be called its structural efficiency, and is a very interesting concept. It measures the extent to which an industry keeps up with the performance of its own best firms. It is a measure of what is natural to call the structural efficiency of an industry - or the extent to which its firms are of optimum size, to which its high cost-firms are squeezed out or reformed, to which production is optimally allocated between firms in the short run (p.262).

If one replaces in the above citation the word *industry* with the word *firm* and the word *firm* with the word *process*, it is clear that the issues arising in structural efficiency measurement for an industry are the same as those arising at the level of the firm when one wants to aggregate the efficiency of its processes.

In reviewing all this material, we discovered astonishing similarities between NDEA models and the forgotten contributions of Kantorovich, Koopmans and Johansen (KKJ). These authors were the first to explicitly state the problem of the efficient allocation of scarce resources in order to maximize production. These initial contributions are strictly connected with the early development of linear programming and the methods of solutions associated with the simplex method. The similarity goes beyond the fact that all these models are using linear programming. If one were to judge this literature in terms of its contribution to optimization theory, then there would be no much originality. To the optimization methodologist, there is nothing really new in any of these contribution, since, from a mathematical perspective, once you write down a linear program that is it. If the reader decides to apply the optimization theorist point of view to this field, then she can stop reading here. On the contrary, we think that there is an original contribution also in the writing and interpretation itself of the linear program at hand because this involves its connection to policy making. In this respect, the contribution of KKJ is substantial and the fact that it has been basically ignored by modern researchers in productivity analysis represents a great disservice to the broader scientific community. In particular, KKJ are using linear programming to give a mathematical and computational representation to policy problems associated with the optimal allocation of scarce resources in order to maximize output. These early authors had clearly in mind a system or network perspective in their approach. These early contributions were sophisticated enough to provide the basis for most of the system efficiency analysis that could be conducted on a modern dataset. They also provided a stringent economic and engineering interpretation of the model that could have formed the basis for a rich analysis. The fact that in the '70s, '80s and '90s these contributions were basically ignored, means that authors started to develop the same model again in the last 10 to 20 years, with the explosion associated with NDEA that we observed in the last 10 years. The reasons why this happened are certainly complex, but a great deal of the explanation may come from the fact that economic, social and cultural thinking in those three decades switched the attention from central planning and government intervention toward a more granular view of society. Accordingly, productivity analysis switched the attention from a system perspective toward a more micro-approach, with an extreme focus on the measurement of efficiency and productivity at the firm level. The complexity of the methodologies associated with the measurement of firm level efficiency has grown in time to an incredible level of sophistication. This sophistication required the simplification of the object of study, and therefore, those early contribution that could have provided the bridge toward a more realistic system analysis have been basically disregarded in favor of a simpler object of inference. The best way of describing this forgotten early literature is to look at the citation count. For the sake of simplicity, we may consider Charnes et al. (1978) (CCR) and Banker et al. (1984) (BCC) the founding papers of DEA analysis and Aigner et al. (1977) the founding paper of stochastic frontier analysis (SFA). DEA and SFA represent the two main approaches to firm level efficiency analysis. These papers received respectively 37,556 citations (Charnes et al.,1978), 21,228 citations (Banker et al. 1984) and 13,213 citations (Aigner et al., 1977). Compare this with the citation count of KKJ. Kantorovich (1939) was published in English in Kantorovich (1960) and it received 990 citations. Koopmans (1953) published on the American Economic Review received 19 citations. The book on which this paper is based (Koopmans 1951) received 1,638 citations. Johansen (1972) book received 633 citations. Charnes and Cooper (1962) (32 citations) knew Kantorovich's and Koopmans' contributions, yet they were very critical of Kantorovich's contribution, focusing their critic on methodological grounds (the reader should notice that any computational and methodological issue was relegated by Kantorovich in an appendix). The Sveriges Riksbank prize committee clearly disagreed with Charnes and Cooper (1962) when assigning the Nobel Prize in Economics to Kantorovich and Koopmans for their contributions to the optimal allocation of scarce resources. This is in line with the reviews of Gardner (1990) and Isbell and Marlow (1961) that stress the importance of Kantorovich's contribution. It is a pity that Johansen was not included in the list of the prize recipients. Johansen's contribution to productivity analysis is in some respects even more important than Kantorovich and Koopmans, in the sense that Johansen was basically proposing to use the KKJ model (based on linear programming) as the tool to be used in the definition of a macro- or aggregate production function based on firm level or microdata on production. Johansen has a clear understanding of the use of such a tool for the micro-foundation of the aggregate production function.

Given that these early contributions are at risk of been completely forgotten by the modern researcher, we decided to organize our story by starting with the analysis of the KKJ model. We then make a leap forward from 1972 to basically 2000, when Fare and Grosskopf (2000) re-introduced a special case of the KKJ model naming it Network DEA. In the 30 years, from 1972 to 2001, nothing really happened in the system approach to productivity analysis except for the fact that researchers actively involved in this field provided a massive amount of methodological machinery for the estimation of firm level efficiency. Even theoretical work on production efficiency mostly focused on the "black box" approach. To be clear, we are not claiming that these 30 years were not useful. We are claiming that they did not advance the research agenda on the system perspective of productivity analysis, which is mostly based on the idea of efficiently allocating scarce resources. Hopefully, we are persuasive enough to show that there are still some quite big challenges in the system approach that are worth more attention than developing another 8 components stochastic frontier model.

The chapter is organized as follows: in section The Origins of Network DEA (1939–1975), we provide a description of the early contributions of Kantorovich, Koopmans and Johansen; in section Shephard, Farrell and the "Black Box" Technology (1977–1999), we very briefly describe the methodological development that happened in the years 1977–1999, by stressing the underlying common "black box" production approach; in section Rediscovery of KKJ (2000–2020), we describe recent developments in 3 apparently disconnected pieces of literature: Network DEA, multi-level or hierarchical models and allocability models; in section Topics for Future Research, we provide a summary of open problems that have not been addressed. Section Epilogo concludes.

The Origins of Network DEA (1939–1975)

In three separate and independent contributions, Kantorovich (1939), Koopmans (1951) and Johansen (1972) laid the foundation for the analysis of efficiency and productivity from a system perspective. Reading these early papers requires some imaginative effort, since the mathematical notation and the language are different from what we use today. The underlying mathematical object is nevertheless the same; therefore, it is just a matter of executing a good "translation". We start this section by describing the model of Kantorovich and introduce the notation in this subsection. As it should result clear by the end of this section, Kantorovich proposed efficiency measurement in a system perspective without making explicit use of intermediate materials and under either variable or non-increasing returns to scale. In view of this fact, the major contribution of Koopmans (1951) is to explicitly account for the use of intermediate materials under constant returns to scale. The introduction of intermediate materials clearly makes the model more flexible and general. Johansen is included in this review because he proposed the same model of Kantorovich under variable returns to scale. Although the model is the same, Johansen interpretation of the model is strikingly different, since Johansen chief interest was in the micro-foundation of the short-run and long-run production function. Of course, it is impossible to make justice to all the details contained in these early papers and they should really be considered the classics of efficiency and productivity analysis that every researcher or practitioner in the field should read carefully. For example, Koopmans' reduction of technology by elimination of intermediate materials has been subsequently used and rediscovered independently by Pasinetti (1973) to introduce the notion of a vertically integrated sector when using input-output tables. We should leave such details out of our review and only focus on the part that concerns the analysis of the production system efficiency.

Kantorovich (1939)

In 1939, Kantorovich presented a research paper (in Russian) proposing a number of mathematical models (and solution methods in the appendix) to solve problems associated with planning and organization of production. The aim of the paper was to help the Soviet centrally planned economy to reach efficiency in production by allocating resources efficiently. Kantorovich's paper was published in English for the first time in 1960 in Management Science (Kantorovich, 1960), and we will refer to the English version of the paper due to our inability to read Russian, although we will refer to it as Kantorovich (1939). Kantorovich introduces his more complicated model (Problem C) in steps by first introducing two more basic models (Problem A and Problem B). In problem A, Kantorovich considers $p = 1, \ldots, P$ machines each one producing $m = 1, \ldots, M$ products. In problem A, the M outputs are produced non-jointly and each machine is used for a specified amount of time in the production of the single product m. This information can be collected in the following data matrix:

$$\mathbf{Y} = \left[y_{mp}\right]$$

where y_{mp} is the quantity of product m that can be produced with machine p in a given reference unit of time. If a machine specializes in the production of a subset of the products, then the coefficients associated with the other products will be equal to zero. It should be noted that in modern terms we would call $\mathbf{Y}$ a data matrix, but we can infer, by the wording Kantorovich is using, that this may just be information on the use of the machines that is obtained via consultation with engineers. Viewing the $\mathbf{Y}$ matrix as a sample is somehow more restrictive than what these early authors had in mind. In general, the information can even come from a booklet of instruction associated with each machine. Kantorovich states his first planning problem in the following way:

$$\begin{aligned} &\max_{\theta,\lambda_{mp}} \theta \\ &st \quad \theta g_m \leq \sum\nolimits_p \lambda_{mp} y_{mp}\,, \ \forall m \\ &\qquad \sum\nolimits_m \lambda_{mp} = 1\,, \ \forall p \\ &\qquad \lambda_{mp} \geq 0 \end{aligned} \tag{4.1}$$

In this formulation $\sum_p \lambda_{mp} y_{mp}$ is the overall amount produced of output m (by all machines jointly) and the coefficients g_m are given and used to determine the mix of the overall output vector produced. Maximizing θ implies that the overall production is maximized in the given proportions g_m. The constraint on the intensity variables λ_{mp} summing up to one is interpreted by Kantorovich as imposing that all machines must be used the whole time (λ_{mp} is the amount of time machine p is used in the production of product m). In modern terms, this constraint has been interpreted as a variable returns to scale constraint (Banker et al., 1984), although the authors proposing such an interpretation don't make any mention of Kantorovich's work. The overall meaning of problem A is to give the maximal production possible (in the given composition g_m) by using all machines at their full capacity level (fully loaded). Later on, in his book, Kantorovich (1965) relaxes this constraint to $\sum_p \lambda_{pm} \leq 1$, therefore allowing for partial use or shut down of machines. The reason for relaxing this constraint is due to the fact that Kantorovich discusses in the book problems associated with capital accumulation. This means that if used for intertemporal analysis, some machines may become economically obsolete if there are other factors that are limiting production. To the best of our knowledge, the use of the model for an analysis of depreciation of capital is still to be implemented along the lines suggested by Kantorovich. In more recent years, this constraint has been interpreted as a non-increasing returns to scale constraint. Because of the special setting of this problem, we want to delve a little bit more into potential interpretations from our point of view (Kantorovich gives several examples of practical problems that can be solved with this model and some of them are astonishingly relevant even today). In particular, if we interpret the *P* machines as being separate production processes, problem A is, in actual fact, a parallel production network, with a linear output set and free disposability of outputs and without inputs (in the basic model Kantorovich assumed that inputs such as energy or labor are available in the right quantities). In particular, this setting allows for the different *P* processes to specialize on different subsets of products, or for them to be just alternative methods of production of the same set of goods. This is in line with the modern approach to Network DEA. Each machine can be allocated to single line production processes, and the only limiting factor is the amount of time the machine can be used for. This means that the output set is linear and problem A can also be interpreted as a basic trade problem where each machine is specializing on the production of the good (or sub-set of goods) for which it has a comparative advantage. The connection with the comparative advantage idea went unnoticed as well, unfortunately, but it is the basis on which one can claim that in general if production units cooperate (or trade if they are in a complete free market) they can yield a bigger output. As a final note, we like to point out that the first constraint in the problem has been stated as an inequality constraint. Strictly speaking, Kantorovich uses an equality constraint, although he mentions that one could allow for "unused surpluses" of the products. Since this is basically a statement of free disposability of outputs, we prefer to state the constraint in its free disposability form.

In problem A, Kantorovich does not make any mention of inputs in the production process and only focuses on a given number of machines and their optimal use in producing given outputs. In problem B, Kantorovich introduces the use of inputs by including information on the use of each possible input (only the one input case is presented in the mathematical problem of Kantorovich's paper, with a mention that extension to other factors is easy and left to the production engineers). In the given reference period of time of use, machine p will be using a given quantity x_{mp} of input (say energy, to follow Kantorovich's example) in order to produce y_{mp} quantity of output m. Generalizing this on the lines proposed by Kantorovich, if the production process uses $n = 1, \ldots, N$ inputs, then x_{nmp} is the quantity of input n used by machine p to produce the quantity of output y_{mp}. If the overall quantity of input n available for production is given by χ_n (notice that this can be equal to the observed overall quantity in the system, or it can be some other quantity set by the researcher), then problem B is:

$$\begin{aligned} &\max_{\theta, \lambda_{mp}} \theta \\ &st \quad \theta g_m \leq \textstyle\sum_p \lambda_{mp} y_{mp}, \ \forall m \\ &\qquad \textstyle\sum_p \sum_m \lambda_{mp} x_{nmp} \leq \chi_n, \ \forall n \\ &\qquad \textstyle\sum_m \lambda_{mp} = 1, \ \forall p \\ &\qquad \lambda_{mp} \geq 0 \end{aligned} \tag{4.2}$$

The second constraint on the overall use of inputs means that the inputs can be a limitational factor for the production of the outputs. Since inputs may be specific to the use of some of the machines, this also means that inputs that are specific to the production of some outputs (output-specific inputs) can be accommodated with Kantorovich problem B. This line of reasoning was proposed recently in Cherchye et al., (2013). One limitation of problems A and B is given by the fact that no joint production of outputs is allowed: each machine is dedicated to the production of a single product at any given time and the overall time for which the machine is available can be allocated to the production of different products. Kantorovich tackles joint production in problem C (which he deems being the most difficult and general). In this problem, each machine p has available $j = 1, \ldots, J$ alternative methods of production for the joint production of the output vector. Therefore, in the given reference time period, machine p can use method of production j to produce the following vector of output quantities $\left(y_{1pj}, \ldots, y_{Mpj}\right)^T$ jointly. Clearly, problems A and B can be embedded as special cases of this more general model by setting $J = M$ and allowing the Υ matrix to be diagonal.

Problem C is stated by Kantorovich as follows:

$$\begin{aligned} &\max_{\theta,\lambda_{pj}} \theta \\ &st \quad \theta g_m \leq \textstyle\sum_p \sum_j \lambda_{pj} y_{mpj} \ , \ \forall m \\ &\qquad \textstyle\sum_p \sum_j \lambda_{pj} x_{npj} \leq \chi_n \ , \ \forall n \\ &\qquad \textstyle\sum_j \lambda_{pj} = q_p \ , \ \forall p \\ &\qquad \lambda_{pj} \geq 0 \end{aligned} \tag{4.3}$$

In problem C of Kantorovich, the activation levels λ_{pj} represent the "quantity of time" each machine p is used with production method j to produce the outputs jointly. Since each method of production j can produce different mixes of outputs, the single line production process can be embedded into this problem as a special case by selecting appropriate methods of production (i.e., one can list the single production line as an additional method of production). Kantorovich does not state explicitly the third constraint on the use of inputs, but by the way the problems are stated, it is clear that this was the intention. Problem C of Kantorovich tackles joint production in the sense that inputs are allocated to machines that can produce joint products.

Since Kantorovich uses in the book the weaker constraint that allows for partial use or shut down of machines, the overall system proposed by Kantorovich can be stated in terms of either variable returns to scale (VRS) or non-increasing returns to scale (NIRS). To the best of our knowledge, Kantorovich never mentioned the assumption of constant returns to scale. On page 375, he states: "*Let there be n machines (or groups of machines) on which there can be turned out m different kinds of output*". "Groups of machines"? If we allow to have replicates of a given machine (let's say we have 100 machines of a given vintage), then this would sum up to an assumption of replicability and we know that replicability together with the NIRS constraint (i.e., divisibility) implies constant returns to scale (CRS). Probably, Kantorovich did not have in mind CRS itself, but rather he was interested in the medium-term output (Soviet Union had 5 years production plans) in a situation where the number of machines is given. In his book later on, he talks about investment and the increase in the production capacity of the economy. Therefore, even if Kantorovich did not have in mind specifically CRS, he was aware of the limitational nature of replicability in the short or medium term and the necessity to deal with expansion in the long term. All in all, one could say that Kantorovich went really close to a notion of CRS by listing the divisibility and replicability assumption. He clearly did not use the axiomatic language that became dominant in the profession later on, but he clearly had in mind these notions and was using them in his examples. In the opening example on page 369 (Table I), Kantorovich gives a clear account of having more than one machine using the same set of technological coefficients. This is a clear cut case of what he means by “groups of machines”: those are replicates of the same machine, i.e., a given number of the same model of machine. Kantorovich gives this idea again in a more general setting on page 385 when he talks about the “Optimum Distribution of Arable Land”. Here, p indexes the different lots of land and each lot can have a different size q_p. Since each lot of land varies in its size, the solution proposed by Kantorovich is equivalent to the constraint $\sum_j \lambda_{pj} = q_p$ which implies that each lot of land needs to be used fully. According to Kantorovich, the q_p are either a natural number representing the number of replicates of machine p, or the size of the lot of land therefore a set of fixed real numbers. There is no account in the paper that makes one think that these fixed numbers can be regarded as decision variables in the optimization problem. If one were to assume them as non-negative decision variables on the real line, then this would sum up to a CRS assumption, but such an assumption is not explicitly stated. In the book, he proposed to relax the constraint to a lower inequality constraint that allows for partial use of the machine. This would amount to the following program:

$$
\begin{aligned}
&\max_{\theta,\lambda_{pj}} \theta \\
&st \quad \theta g_m \leq \sum_p \sum_j \lambda_{pj} y_{mpj}\ , \ \forall m \\
&\qquad \sum_p \sum_j \lambda_{pj} x_{npj} \leq X_n\ , \ \forall n \\
&\qquad \sum_j \lambda_{pj} \leq q_p\ , \ \forall p \\
&\qquad \lambda_{pj} \geq 0
\end{aligned}
\tag{4.4}
$$

What can we say in terms of interpretation of the Kantorovich model? The first point to make clear is that the model has two levels of decision making in problem C. One can easily grasp that the intensity variables λ_{pj} depend both on the machine used and on the selected method of production. Now, if we rename “machines” as “processes” and “methods of production” as “firms”, in all effects we have a model which is producing M outputs, using N inputs and each firm j is using P production processes to accomplish this production. This is the very first example of an attempt to open the black box of production, even before the black box of production idea was proposed. Kantorovich's model is a fully fledged parallel production network under alternative specifications of returns to scale.

At this point, we should also notice that the data structure that Kantorovich had in mind is three dimensional. By looking at the input data, we have P matrices $\mathbf{X}_p$ where the inputs are listed in the rows and the production methods in the columns. If we overlap all these matrices, we obtain a three-dimensional data structure:

We shall see in the next subsection that Koopmans (1951) is using the same data structure by stacking these matrices into a large two-dimensional matrix. Kantorovich does not discuss explicitly how many replicates of each machine we should use, but if we were to assume a long-term view and make the number of replicates a variable, then we could solve the previous problem for several values of q_p and choose the ones that maximize production for the given level of inputs available. This would make the number of "firms" in the industry a variable of choice like in Ray and Hu (1997) or Peyrache (2013, 2015). Moreover, the model also includes output-specific inputs (Cherchye et al., 2013) by designing the data $\left(y_{mpj}, x_{npj}\right)$ appropriately in order to make them specific to some of the processes.

If we account for the fact that this paper was published in Russian in 1939 and in English in 1960, this means that many production models recently proposed in the literature can be embedded as special cases of Kantorovich model and have been floating around for at least 60 years. The bottom line of this analysis is that in Kantorovich modeling J is the number of methods of production (this can be observed firms) and P is the entities we are evaluating. The coefficients $\left(y_{mpj}, x_{npj}\right)$ will determine the particular interpretation we want. Therefore, we can also obtain the widely celebrated output-oriented DEA models under VRS, NIRS (or CRS if we include replicability of the machines) by setting $P = 1$ and $\left(y_{mpj}, x_{npj}\right) = \left(y_{mj}, x_{nj}\right)$ where the dependence on the process has been dropped in the notation because $P = 1$ and one is evaluating the efficiency of the production plan $(\mathbf{y}_0, \mathbf{x}_0)$. Output orientation is obtained as a special case by setting $g_m = y_{0m}$. In fact, this is even more general than the output-oriented model because the projection is dictated by the g_m coefficients. One is left to wonder if the 37,000 citations of the CCR model or the 21,000 citations of the BCC model are better deserved than the less than 1,000 citations of Kantorovich's work, especially considering the exponential growth in Network DEA that we observed over the past 5–10 years.

The chief interest of Kantorovich is into optimal allocation of resources in order to maximize the output of the system. He does not show any interest in the efficiency at a more granular level and he takes for granted that if a machine is not used efficiently then it should be used at the efficient level (this is implicit in the formulation of the problem). Since the objective function is maximizing the overall output produced, this corresponds to an industry model where firms have a network production structure and the production runs in parallel without any flow of intermediate materials from one process to another. The words of Kantorovich himself are better than any explanation:

> There are two ways of increasing the efficiency of the work of a shop, an enterprise, or a whole branch of industry. One way is by various improvements in technology; that is, new attachments for individual machines, changes in technological processes, and the discovery of new, better kinds of raw materials. The other way - thus far much less used - is improvement in the organization of planning and production. Here are included, for instance, such questions as the distribution of work among individual machines of the enterprise or among mechanisms, the correct distribution of orders among enterprises, the correct distribution of raw materials, fuel, and other factors. (p. 367)
>
> ... I discovered that a whole range of problems of the most diverse character relating to the scientific organization of production (questions of the optimum distribution of the work of machines and mechanisms, the minimization of scrap, the best utilization of raw materials and local materials, fuel, transportation, and so on) lead to the formulation of a single group of mathematical problems.
>
> I want to emphasize again that the greater part of the problems of which I shall speak, relating to the organization and planning of production, are connected specifically with the Soviet system of economy and in the majority of cases do not arise in the economy of a capitalist society. There the choice of output is determined not by the plan but by the interests and profits of individual capitalists. The owner of the enterprise chooses for production those goods which at a given moment have the highest price, can most easily be sold, and therefore give the largest profit. The raw material used is not that of which there are huge supplies in the

> country, but that which the entrepreneur can buy most cheaply. The question of the maximum utilization of equipment is not raised; in any case, the majority of enterprises work at half capacity.
>
> Next I want to indicate the significance of this problem for the cooperation between enterprises. In the example used above of producing two parts (Section I), we found different relationships between the output of products on different machines. It may happen that in one enterprise, A, it is necessary to make such a number of the second part or the relationship of the machines available is such that the automatic machine, on which it is most advantageous to produce the second part, must be loaded partially with the first part. On the other hand, in a second enterprise, B, it may be necessary to load the turret lather partially with the second part, even though this machine is most productive in turning out the first part. Then it is clearly advantageous for these plants to cooperate in such a way that some output of the first part is transferred from plant A to plant B, and some output of the second part is transferred from plant B to plant A. In a simple case these questions are decided in an elementary way, but in a complex case the question of when it is advantageous for plants to co-operate and how they should do so can be solved exactly on the basis of our method.

This is an incredibly fascinating sentence in all respects, but Kantorovich goes on:

> The distribution of the plan of a given combine among different enterprises is the same sort of problem. It is possible to increase the output of a product significantly if this distribution is made correctly; that is, if we assign to each enterprise those items which are most suitable to its equipment. This is of course generally known and recognized, but is usually pronounced without any precise indications as to how to resolve the question of what equipment is most suitable for the given item. As long as there are adequate data, our methods will give a definite procedure for the exact resolution of such questions. (p. 366, Kantorovich, 1939).

This is a clear statement and description of what we would call today an industry model, centralized allocation model or network model. Moreover, the statement is so clear (and does not involve formulas) that makes one wonder why we write the same sort of problems in a much more intrigued and cryptic fashion. Kantorovich goes on and discusses: optimal utilization of machinery, maximum utilization of a complex raw material, most rational utilization of fuel, optimum fullfilment of a construction plan with given construction materials, optimum distribution of arable land and best plan of freight shipments. Only a researcher fixated with finding the next generation of complicated models that will deliver improbable estimates of individual firm efficiencies could deny the practical and empirical relevance of these problems for the modern economy, half of which is run with centrally planned operations and the other half is regulated to solve some sort of market failure.

Kantorovich's work was a major breakthrough in productivity and efficiency analysis. The solution methods for the associated linear programs developed around the same time by Dantzing in the west resulted to be more powerful. But from the perspective of organizing an economy, sector, industry or company in the best possible way (which is at the end the core of productivity analysis), Kantorovich's contribution stands as being the most significant contribution of the last 80 years. It lays clearly the foundation for work related to optimal allocation of resources in order to maximize system output. In fact, computational issues are relegated by Kantorovich into an appendix. It is somehow puzzling that Charnes and Cooper (1962) were so critical of Kantorovich's work and were focusing almost exclusively on the computational aspects rather than looking into the ways that the model could be used for empirical analysis and policy making. Johansen (1976) and Koopmans (1960) clearly recognize the importance of Kantorovich's work. The "critique" of Charnes and Cooper (1962) is even more astonishing considering that some of the models proposed by these authors later on were actually embedded as special cases of Kantorovich's model. Given the influence of the CCR and BCC models in efficiency analysis, it would have made sense to include Kantorovich work as one of the seminal papers that introduced a more intriguing production structure. In fact, Koopmans (1960) words on Kantorovich's work are the best way of describing the importance of this contribution:

> The application of problems "A", "B" and "C" envisaged by the author include assignment of items or tasks to machines in metalworking, in the plywood industry, and in earth moving; trimming problems of sheet metal, lumber, paper, etc.; oil refinery operations; allocation of fuels to different uses; allocation of land to crops, and of transportation equipment to freight flows. One does not need to concur in the authors' introductory remarks comparing the operation of the Soviet and capitalist systems to see that

> the wide range of applications perceived by the author make his paper an early classic in the science of management under any economic system. For instance, the concluding discussion anticipating objections to the methods of linear programming has a flavor independent of time and place.
>
> There is little in either the Soviet or the Western literature in management, planning, or economics available in 1939, that could have served as a source for the ideas in this paper, in the concrete form in which they were presented. From its own internal evidence, the paper stands as a highly original contribution of the mathematical mind to problems which few at that time would have perceived as mathematical in nature - on a par with the earlier work of von Neumann on the proportional economic growth in a competitive market economy, and the later work of Dantzing well know to the readers of Management Science.

The Nobel Prize committee clearly listened to Koopmans' words when assigning the 1975 economic prize to both of them for their major contribution in the science of the optimal allocation of scarce resources.

Koopmans

Kantorovich's examples always involve one particular industry or a particular group of machines. In his 1965 book, there is a more general discussion on how one could potentially extend these ideas to the whole economy as well. As we shall see in this subsection, from the point of view of system efficiency, Koopmans' most important contribution was to actually provide a way of measuring efficiency for the whole economy, by taking into explicit account the use and flows of intermediate materials across the different nodes of the network (the different sectors or activities of the economy). In 1951, Koopmans collected the proceeding of a conference in a book titled "Activity analysis of Production and Allocation". In the opening statement of the book, Koopmans states:

> The contributions to this book are devoted, directly or indirectly, to various aspects of a fundamental problem of normative economics: the best allocation of limited means toward desired ends.

There are various ways of presenting Koopmans' contribution. The way we want to approach the presentation here is to have it in connection with the model of Kantorovich. Although the paper of Kantorovich was not known to Koopmans in 1951 (therefore Koopmans' contribution is completely independent from Kantorovich's contribution), the two papers approach the same empirical problem using very similar methods. Therefore, we see the two contributions as complementary rather than competing with each other.

As noted in Charnes and Cooper (1962), Kantorovich is ambiguous about the sign of the data. Quite in stark contrast, Koopmans is very clear about the underlying conditions under which the "efficient production set" is non-empty and this is a necessary condition for the model presented by Kantorovich to have a basic feasible solution. Koopmans presents all his results under the CRS assumption (although he mentions that CRS is not necessary and results can be generalized to variable returns to scale). If we make the coefficients q_p free non-negative decision variables in problem (4.3), then the intensity constraint $\sum_p \lambda_{pj} = q_p$ is redundant and we can omit it (which is the equivalent to assume CRS). Before we proceed and write the model explicitly, it is useful to provide the classification of inputs and outputs proposed by Koopmans. Koopmans uses the same matrices of data for the inputs and the outputs, but he introduces an additional set of matrices, which are the matrices of intermediate materials. We will indicate intermediate products as z_{lpj} with $l = 1, \ldots, L$. While Koopmans assumes that all input and output quantities are positive, the L intermediate materials can be both positive or negative. If z_{lpj} is negative, then it represents the quantity of intermediate l used as an input in process p with production method j. If z_{lpj} is positive, then it represents the quantity of intermediate l produced as an output in process p using method of production j. This is equivalent to adopting a netput notation for the intermediates. In particular, Koopmans is assuming that for each intermediate l, there is at least one process that is using it as an input ($z_{lpj} < 0$ for at least one p and one j) and is produced as an output by at least one process ($z_{lpj} > 0$ for at least one p and one j). If this condition does not hold, then the intermediate should be classified as either an input or an output (depending on its sign). Intermediate materials are produced within the system to be used within the system. Koopmans imposes explicitly that the overall net production of every given intermediate must be non-negative (otherwise production would be impossible because it would require some flow of the intermediate from outside the system), which amounts to adding the following constraint to model (4.3):

$$\sum_p \sum_j \lambda_{pj} z_{lpj} \geq \eta_l \,, \ \ \forall l = 1, \ldots, L \tag{4.5}$$

In actual fact, Koopmans allows this constraints to be tightened by the quantities (η_l), by proposing that some of the intermediate materials may be flowing into the system. In other words, these coefficients allow for situations in which some intermediate materials must be available before starting production, or some intermediate materials must be produced as final outputs to be used in future production. The sign of the η_l coefficients is negative if the intermediate is an input that must be available before starting production, and they are positive if the intermediate must be produced above a certain quantity as a final output. These quantities play the same role here as the overall quantities χ_n in Kantorovich's model. Adding this constraint to problem (4.3) and omitting the intensity variable constraint to allow for CRS, returns the Koopmans' model of production.

Koopmans introduces a more parsimonious way of representing the system and the underlying data of the problem. The best way of introducing such notation is by looking at the stacking of the three-dimensional matrices of Kantorovich. If we stack all the input matrices together and transpose them, we obtain:

$$\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_P] \tag{4.6}$$

Although this makes the notation a bit more confusing, we will refer to $\mathbf{X}_p$ as one particular two-dimensional matrix of inputs for process p as in the representation of Kantorovich. And we will refer to $\mathbf{X}$ as the stacked two-dimensional matrix composed of the stacking of all of the P input matrices. Notice that each row of matrix $\mathbf{X}$ represents now a particular input; that is, the dimension of the matrix is $N \times (J + P)$. We can define in the same way the output matrix

$$\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_P] \tag{4.7}$$

and the matrix of intermediates

$$\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_P] \tag{4.8}$$

We can now stack these large matrices into the following one:

$$\mathbf{A} = \begin{bmatrix} -\mathbf{X} \\ \mathbf{Z} \\ \mathbf{Y} \end{bmatrix} \tag{4.9}$$

In this matrix, each column represents the netput of a given production process. Koopmans calls the columns of this matrix “basic activities”. Notice that if the three-dimensional matrix of Kantorovich is sparse, then Koopmans’ representation provides a more parsimonious way of representing the data, since one can eliminate all the columns that have zero for all inputs and outputs (all columns filled with zeros only). In Koopmans, the technology matrix is dense, while in Kantorovich it could be sparse. On the other hand, if one were to introduce VRS constraints on the intensity variables for all processes, then Kantorovich’s representation is more exhaustive and general, since the processes are accounted for in a more explicit way. To do the same with the more succinct way of Koopmans, one need to introduce an indicator matrix with as many columns as the number of intensity variables and as many rows as the number of processes. This matrix will only contain indicator variables, i.e., zeros and ones. Then, the intensity variable constraints can be represented as:

$$\mathbf{W}\boldsymbol{\lambda} = \mathbf{1}_P \tag{4.10}$$

where $\mathbf{1}_P$ is a column vector of ones of dimension P. If we call $\boldsymbol{\pi}$ a generic $(N + M + L)$ netput vector, then we can obtain the very parsimonious representation of the production possibilities set proposed by Koopmans:

$$\boldsymbol{\pi} = \mathbf{A}\boldsymbol{\lambda}\,,\ \ \boldsymbol{\lambda} \geq \mathbf{0} \tag{4.11}$$

where $\boldsymbol{\lambda}$ has all the λ_{pj} coefficients stacked together. If we call the intensity variables of process p, $\boldsymbol{\lambda}_p = \left[\lambda_{p1}, \ldots, \lambda_{pJ}\right]^T$, then the stacked vector of intensity variables for the system is:

$$\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \vdots \\ \boldsymbol{\lambda}_P \end{bmatrix} \tag{4.12}$$

Although this is a parsimonious representation, Koopmans’ suggestion of introducing limitations on the primary factors of production is better written in formal terms by looking at the individual input, output and intermediate matrices. It should also be stressed that Koopmans' interest is in determining efficient sets and he does not really propose (contrary to Kantorovich) an objective function to determine maximal or optimal production. If we were to choose the same objective function of Kantorovich, then we would write the optimization model as (where we omit non-negativity constraints on the decision variables $\boldsymbol{\lambda} \geq \mathbf{0}$):

$$\begin{aligned} &\max_{\theta,\boldsymbol{\lambda}} \theta \\ &st \quad \theta\mathbf{g} \leq \mathbf{Y}\boldsymbol{\lambda} \\ &\qquad \boldsymbol{\eta} \leq \mathbf{Z}\boldsymbol{\lambda} \\ &\qquad \boldsymbol{\chi} \geq \mathbf{X}\boldsymbol{\lambda} \end{aligned} \tag{4.13}$$

As said earlier, this program is expressed under the assumption of CRS (as in Koopmans). One can introduce VRS by adding the constraint $\mathbf{W}\boldsymbol{\lambda} = \mathbf{1}_P$, or NIRS by adding the constraint $\mathbf{W}\boldsymbol{\lambda} \leq \mathbf{1}_P$. Alternatively, one can take the notion of replicability of Kantorovich and write this constraint as $\mathbf{W}\boldsymbol{\lambda} = \mathbf{q}$ where $\mathbf{q}$ are pre-specified levels of replication. The new explicit constraint on the intermediates states that given the activation levels represented by the intensity variables λ_{pj}, the overall net production of intermediate material l of the system must be non-negative. This means that the system is producing enough intermediate material to satisfy the use of it in all production processes that require it as an input. It should be noted that under CRS the notation is simplified further because there are no restrictions on the λ_{pj}, apart from non-negativity constraints.

What can we say about Koopmans' model in connection with system efficiency? The intelligent reader will convince herself that Koopmans' technology can embed a whole lot of network structures (actually the large majority) that have been produced in the last few years. We shall discuss this briefly in the next few sections, by giving some examples. We should also point out that Koopmans has an explicit discussion on the prices associated with the efficient subset of the production set. This set of prices (which is nothing more that the separating hyperplane at the optimal solution of problem 4.13) is discussed by Koopmans in connection with planning problems that involve decentralized decisions. In this sense, the price vector is used by Koopmans to incentivize individual production units to reach the optimal plan set out by the central planner. Kantorovich (1965) in his book takes up this discussion even in a more explicit way, by suggesting that this set of supporting prices would permit the fulfillment of the 5 year plan, by making the best use of the limited economic resources at hand. Of course, neither Koopmans or Kantorovich introduced the dual problem that would set the optimization problem directly in terms of supporting shadow prices. But both of them had clearly in mind that such a vector of supporting prices could play a key role in practice. Koopmans discussed this explicitly and Kantorovich implicitly by proposing his solution method based on the "resolving multipliers". The issue of decentralization of the plan by providing individual production units with a set level of prices at which they could trade their inputs and outputs has not been used as a tool for implementation of the optimal solution.

All in all, Koopmans' contribution, especially if read in connection with Kantorovich's paper, represents another big leap forward in our ability to represent production systems. The introduction of the CRS assumption and the constraints associated with the use and flows of intermediate materials open up wide possibilities of applications and actually nest many of the current proposals in Network DEA analysis. Although Koopmans' paper is well known within the productivity community (contrary to Kantorovich's paper), his general representation of the technology set that basically includes network models has been widely neglected, with the scientific community posing excessive attention on the definition that Koopmans gives of an efficient set. This is a misplaced interpretation and minimizes the contribution of Koopmans to productivity and efficiency analysis, since the notion of efficiency of Koopmans was already proposed by Pareto. The main point of Koopmans' analysis regards (in line with Kantorovich) the efficient allocation of a limited amount of resources to produce the maximal possible output. His representation of the technology set associated with this problem is so general and simple that puts to shame many modern representations (including the one of the authors, Peyrache and Silva, 2019). Everyone should read Koopmans' book if interested in efficiency and productivity analysis in order to experience that feeling of satisfaction and fulfillment that only the reading (and studying) of the great classical thinkers of our time can provide—a feeling (to say this using Koopmans' words) that "has a flavour independent of time and place".

Johansen

Johansen (1972) had a chief interest in the micro-foundation of the aggregate production function. Johansen's setting of the problem was an aggregation from the firm production function to the industry production function. If we call $f(x)$ the firm production function and there are J firms in the industry, then Johansen defines the industry production function as:

$$F\left(\sum_j x_j\right) = \max_{x_j} \sum_j f(x_j) \tag{4.14}$$

This means that if the overall quantity of input of the industry is $\chi = \sum_j x_j$, then the industry overall maximal production is obtained by allocating the industry input χ to individual firms optimally by choosing the appropriate allocations x_j. Johansen notices that if the firm level production function is approximated by a piece-wise linear envelope of the observed data points, the previous maximization problem becomes a linear program. In fact, the linear program associated with such a specification is the same as in Kantorovich's specification. This is not surprising since the objective of Johansen's problem is to choose the allocation of resources (inputs) to the various firms in a way that maximizes the overall output produced by the industry. Johansen calls this approach the nonparametric approach to the micro-foundation of the aggregate production function. He goes on discussing notions of short-run vs long-run choices, and most importantly, he notices that if one is willing to make additional assumptions on how the inputs are distributed across firms one can make more explicit the parametric form of the aggregate production function. For example, he notices that the contribution of Houthakker (1955) is an example of such an approach: if one assumes that the inputs are distributed as a generalized Pareto, then the aggregate production function is Cobb-Douglas. Interestingly, Houthakker was making an explicit connection to the activity analysis model of Koopmans. This fact has been recently used by Jones (2005) in macroeconomic modeling.

Johansen further discusses issues associated with technical change and how to introduce it into the model. Johansen's book is a source of inspiration for work in productivity analysis that still has to happen. All in all, Johansen is providing an explicit link to economics and he is suggesting a way of proceeding that makes use of the activity analysis model by looking at the distribution of inputs across firms. Interestingly, this did not give rise to a proper research program exploring how to use statistical methods to estimate density functions on data in order to obtain the industry production function. This work is still far from being accomplished, and in this sense, Johansen's (1972) book is an important source of inspiration. This could help the scientific community in efficiency and productivity analysis to make a more explicit connection and build a bridge and a methodology that can be used in macroeconomic modeling. Among the three authors that we reviewed so far, Johansen is definitely extremely original and also the most neglected of the three.

Summing Up: The KKJ (Kantorovich-Koopmans-Johansen) Model

We shall refer to these early contributions as the Kantorovich-Koopmans-Johansen (KKJ) model and consider the specification of program (4.13) with the associated discussion on the constraints on the intensity variables to characterize returns to scale as the benchmark model. This model allows for various forms of returns to scale, and at the same time, it makes use of intermediate materials, therefore making it suitable to represent networks system, where the nodes of the system are connected by the flow of intermediate materials.

Before we close this long section on the KKJ model, it is useful to show its application to some of the current models proposed in the literature, just to give a flavor of the flexibility and generality of the KKJ model. Let us assume for simplicity that there are only two processes, 3 firms (or methods of production), two inputs and two outputs. If the two processes are independent, with input 1 producing output 1 in process 1, and input 2 producing output 2 in process 2, then the associated input and output matrices would be:

$$\mathbf{X} = \begin{bmatrix} x_{111} & x_{112} & x_{113} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{221} & x_{222} & x_{223} \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} y_{111} & y_{112} & y_{113} & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{221} & y_{222} & y_{223} \end{bmatrix}$$

The first 3 columns of these matrices represent process 1, and the second 3 columns process 2. Since input 1 enters with zeros in process 2 and so does output 1, this means that process 1 is producing output 1 using input 1; that is, input 1 is specific to the production of output 1. This is true for process 2 as well. This is an example of two single production lines working in parallel. If we wanted these two production lines to work sequentially in a series two-stage network, then the matrix of intermediates would be:

$$\mathbf{Z} = \begin{bmatrix} z_{111} & z_{112} & z_{113} & z_{123} & z_{123} & z_{123} \end{bmatrix}$$

with the caveat that the first 3 entries of this matrix would be positive (the intermediate material is an output of process 1) and the second 3 entries would be negative (the intermediate material is an input of process 2). This provides the KKJ representation of the widely "celebrated" two-stage Network DEA model. One can easily see that by building these basic matrices in an appropriate manner, it is possible to cover such a wide variety of network structure that we are not even sure any of the current proposals falls out of this representation. For example, the joint inputs model of Cherchye et al. (2013) requires that if an input is provided in a given quantity to one process, then it is available in the same quantity to all other processes (it is a public good). Suppose a third input is available, then we would change the input matrix to:

$$\mathbf{X} = \begin{bmatrix} x_{111} & x_{112} & x_{113} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{221} & x_{222} & x_{223} \\ x_{311} & x_{312} & x_{313} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{311} & x_{312} & x_{313} \end{bmatrix}$$

and as the reader can verify the quantity of input available to process 2 is the same as process 1. Even if rows 3 and 4 represent the same physical input, we separated them so that when summing up the total quantity of input available to the system, these quantities are not double counted. By splitting and creating additional rows and columns and creating fictitious inputs and outputs, one can accommodate so many structures that the only limitation is the creativity and imagination of the applied researcher. This would, for example, allow us to keep the level of the intermediate flows at the observed level, rather than making them change in the optimal solution, de facto nesting so-called fixed link Network DEA models. This can be accomplished by adding a fictitious number of rows to the matrices in order to preserve the current allocation.

Koopmans published his work in 1951, Kantorivich in English in 1960 and Johansen his book in 1972. The Nobel Prize was assigned to Kantorovich and Koopmans in 1975. Therefore, if a martian were to come to planet earth in 1976, she would have been provided with a strong mathematical model to deal with problems associated with the optimal allocation of resources in production systems. It is very likely that the martian would have started to look at issues associated with the use of such a model and the associated collection of data, and she would have delved into a list of issues that we are going to describe at the end of this chapter. But this is not what we have done on planet earth. With the contributions of Charnes et al. (1978), Banker et al. (1984), Aigner et al. (1977), Fare and Lovell (1978) and the associated work on duality theory of Ronald Shephard, the scene was set for studying production using the black box technology approach. To be fair, we should also point to the fact that at that time the available data was more limited and this may have contribute to shift the attention toward firm level analysis. Certainly, the boom in NDEA publications in the last 10 years has partly to do with the availability of more refined datasets that contain information at a lower level of aggregation and actually permit to go beyond black box analysis. Even so, it is puzzling that researchers focused on firm level efficiency, given that a firm level dataset allows at least the possibility of carrying out the industry model analysis so well presented and discussed in Johansen. At the very least, the Johansen model should have had become a basic analytical tool in the efficiency and productivity community.

In any event, starting in the late ’70s for about 30 years, an entire generation of researchers in efficiency and productivity analysis has worked on the basic assumption that input data and output data are available at the firm level and the main focus of the analysis should be the one of measuring the efficiency and productivity of individual firms. This paradigm laid the foundation for all subsequent work on stochastic frontier analysis, DEA, index numbers, economic theory of production and aggregation and duality. Very little if anything has been done during these 30 years in terms of looking “inside” the black box, which was what the KKJ model basically does. By saying this, we don’t want to minimize the impact of what has been done in terms of research in efficiency and productivity analysis. We just want to point out to the fact that in one way or another the memory of the KKJ model has been lost, and a lot of the effort that went into building Network DEA models could have been saved if the KKJ model were to be credited the correct amount of attention and importance in this field of study. In some sense, we lost a lot of the creativity and understanding of how to optimally organize and measure the efficiency of a system of production that these early authors so forcefully and elegantly described. In exchange for it, we greatly simplified the object of our study. After simplifying it, the research problem has been reduced to the measurement of the efficiency of a single individual firm. Starting at the end of the ’70s, the scene was set to research and deliver an impressive methodological machinery that keeps growing at the present day and allows the modern researcher to have very flexible strategies to estimate the black box production technology.

Shephard, Farrell and the “Black Box” Technology (1977–1999)

In two independent contributions, Farrell (1957) and Shephard (1970) laid the foundation for what would become the “black box” technology and the basis of the successive 30 years of research in efficiency and productivity analysis. This is clearly the case if one looks at the citation count of Farrell: with 23,879 citations, this is definitely the founding paper of modern productivity analysis. Shephard’s 1970 book received 4,887, but one should keep in mind that this is a theoretical contribution, and for being a theoretical contribution, this represents a high number of citations. From the perspective of our discussion, the main outcome of these two contributions is to set the scene for a simplified object of inquiry, shifting the attention from the optimal allocation of resources and the associated problems of measurement, toward the optimal use of those resources at the firm level. The firm is considered the basic unit of the analysis, and problems associated with reallocation of inputs and production across production units are rarely taken into consideration. These two contributions formed the basis for successive work on production frontier estimation, inference and theoretical development. The reference to the firm as the basic unit of analysis, without reference to the component production processes or the allocation problems across different firms, has given rise to the definition of such an approach as a “black box” approach. The firm is a “black box” in the sense that we only observe the inputs that are entering production and the outputs that are exiting as products, but we do not observe what happens inside the firm. This is in sharp contrast to both the KKJ approach and the Network DEA approach.

The best way of describing this is to look once again at citation count as a rough measure of the popularity of the main contributions in the field. Aigner et al. (1977) and Meeusen and van Den Broeck (1977) received respectively 13,229 and 7,811 citations, laying the foundation for the research program on stochastic frontier production function estimation and inference. Subsequent work (continuing today) made the model more and more flexible considering issues associated with functional form specification, panel data, additional error components and all the methodological machinery that is still under development, providing a large body of models and methods for estimation and inference. Charnes et al. (1978) and Banker et al. (1984) (after renaming the linear activity analysis model DEA) received respectively 37,581 and 21,240 citations, setting the agenda for research in DEA and estimation of production frontiers and technical efficiency at the firm level. This stream of literature saw the development of a plethora of efficiency measures (radial, slack based, directional, etc.) and alternative ways of specifying returns to scale, and relaxation of the convexity assumption. Fare and Lovell (1978), with a citation count of 1,459 (high for a theoretical contribution), made the connection between economic theory, duality and efficiency and productivity analysis; subsequent work will see the Shephard duality approach extended to various alternative notions of technical, cost, revenue and profit efficiency.

All of those contributions have a commonality in the fact that they are based on the black box technology and they lack any interest in the problems of allocation of resources that was the core of the early development of the KKJ model. Therefore, the subsequent work in efficiency analysis, at least until the first decade of this century, basically "forgot" the problem of optimal allocation of resources and took the route of simplifying the policy problem to the analysis of the firm and its efficiency in various forms. By no means, we are implying that this work was not useful: quite on the contrary, this work equipped the modern researcher with a tremendous set of tools to analyze firm level dataset and the various measures of efficiency associated with the black box technology idea. The side effect of this massive amount of work that went into estimation, inference and theoretical development of the black box technology is that the latest generation of researchers in productivity analysis has no memory of the early developments associated with the KKJ model. Starting with the contribution of Fare and Grosskopf (2000), the field started re-discovering the problem of optimal resource allocation, without the knowledge of the work of the KKJ model.

Rediscovery of KKJ (2000–2020)

The literature on system efficiency has grown disperse, and in fact, the name "production system" is rarely used. Instead, there are the following strands of the efficiency measurement literature that can be considered within this production system perspective:

- Netowork DEA models;
- Multi-Level or Hierarchical models;
- Input-output allocability models.

It should be noted that in the literature we found a variety of names trying to describe the same sort of problems—for example, "industry models" have also been called "centralized allocation models". The rationale we follow for our classification is based mainly on the separation between the decision problem of allocating resources to the different nodes of the system, from the efficient use of these resources in production. In Network DEA models, the focus is typically oriented toward the firm and its internal structure. Clearly, there are two layers of decision making here, and in this sense, these models could also be discussed under the multi-level models. We keep Network DEA models separated from the rest because of the large strand of the literature dealing with the internal structure of the firm. In multi-level models, there are various layers of decision making delivering the observed allocation of resources. In fact, in such a system, decision making happens at all the various levels: at the level of the production process, at the level of the firm and at the level of the industry or the economy as a whole (we include industry models in this class). When studying production system models, it is important to categorize the types of inputs and outputs that are used and produced. The literature has, most often than not, ignored this classification, except for certain cases where explicitly some inputs are considered allocatable and the optimal allocation is to be determined; or some cases where the specificity of some inputs in the production of only one or a subset of outputs is considered. As a result, we also consider this strand of literature separated from the rest because it explicitly deals with the definition itself of inputs and outputs. We call this stream "Input-output allocability models". Note that this division or classification is arbitrary, as indeed are all classifications that can be found in the literature. This may be confusing, and in a sense, this could be one of the reasons why these different streams of literature are developing independently.

The KKJ model looks at optimality conditions for the system as a whole: the associated efficiency measure is computed for the whole system, being it an industry, a firm or the whole economy. One of the merits of the last 20 years of research on this topic has stressed the importance of assigning the overall inefficiency of the system to the different components. We shall not discuss these contributions in too much detail because that would be out of the scope of this chapter and would take excessive space. One could even make the argument that assigning efficiency to the different components of the system is not really useful, since the KKJ model is already providing targets for the different components that would make the whole system efficient. We rather focus on the connection between the KKJ model and this recent literature in terms of the structure of the underlying system.

In what follows, we will explain what each of the aforementioned strands of the literature aims to do in terms of efficiency measurement and we will explain how these various strands are in fact interconnected (and how they relate to the KKJ model). As a matter of fact, the relationship between the various strands of the literature is hardly acknowledged in the literature.

Network Models

Many network models (in particular those that do not allow for intermediate materials) are in all aspects similar to industry models, but authors have not recognized this link. This has happened mainly because the two types of analysis have somehow different objectives. Whereas in the multi-level model literature, it has been recognized that the aggregate is more than the sum of its parts because of allocation inefficiencies, in network models, most often than not, allocation issues are not even mentioned and the problem is mainly mathematical: that of providing an efficiency of the parts and of the whole and aggregating the parts to form the whole or disaggregating the whole into its parts. In this mathematical exercise, authors have missed the most important issue: that the whole is different from the sum of its parts and possesses characteristics that parts do not. In particular, as we saw with the KKJ model, allocation inefficiencies are somehow the core of this type of analysis.

As mentioned in the introduction, the Network DEA literature is growing at a very fast pace. Kao (2014) provides a review of Network DEA models and includes them into 7 types (basic two-stage structure, general two-stage structure, series structure, parallel structure, mixed structure, hierarchical structure and dynamic structure). It is interesting to verify that more than 170 studies in his Table 1 (pages 11 and 12) are two-stage models (representing more than 50% of the total number of studies). Another important remark is that allocation issues are not addressed in this literature. In the same year, Castelli and Pesenti (2014) also reviewed the Network DEA literature and classified papers into 3 categories: Network DEA; shared flow models; and multi-level models. Interestingly, Castelli and Pesenti (2014) claim that in Network DEA the subunits do not have the ability to allocate resources, and therefore, they assume that when this assumption is dropped models fall into the shared flow models (which are essentially network models where allocation of resources is allowed). Castelli and Pesenti (2014) basically recognized the fact that most network models are ignoring the resource allocation issue and solve the problem by assuming that the word "network" is unrelated with resource allocation issues. In addition, Castelli and Pesenti (2014) interpret dynamic models as network models, and therefore, no reallocation of resources is allowed. On the contrary, Kao (2014) considers dynamic models as a separate type of network model. Dynamic models have, indeed, been treated as a separate type of network models as the review by Fallah-Fini et al. (2014) testifies. In this review, the authors distinguish between alternative dynamic models by the way intertemporal dependencies are treated (as production delays, as inventories, as capital related variables, as adjustment costs and as incremental improvement and learning processes). Agrell et al. (2013) also reviewed series or supply chain network models in depth, pointing out the prevalence of two-stage network models and the fact that "most models lack a clear economic or technical motivation for the intermediate measures" (p. 581).

To the best of our knowledge, the term "Network DEA" was introduced in the literature with the work of Fare and Grosskopf (2000). This work is a follow up of Fare (1986), where dynamic models have been modeled as a network structure for the first time. In these models, a firm observed in different periods of time is analyzed as a whole entity since it is assumed that certain factors pass from period to period and work as a link between time periods. This means that the same firm in different time periods should be assessed as a whole entity, or as a system that is temporally interconnected. Clearly, this bears connections with supply chain or series models where some factors flow from one process to the next as intermediate factors. Therefore, dynamic models can be seen as series models of network structures. In what follows, we start by presenting dynamic network models. This choice is dictated by the fact that dynamic network models can be viewed as the more general class, of which the series and parallel network structures are special cases. This choice will also make the connection to the KKJ model more clear.

Dynamic Network DEA Models

Fare (1986) proposed models with separate reference technologies for each time period. The author classifies inputs into two categories: (i) inputs that are observed and allocated to each time period and (ii) inputs whose total amount (across all time periods) is given, but not its time allocation. The second class of inputs is also considered in some multi-level models, where the allocation of some inputs is not observed. Fare and Grosskopf (1996) take up on this work, and introduce the idea of intermediate factors linking time periods. This idea is at the basis of most series network models (dynamic or not).

One of the first models to be employed for dynamic network models was that of Fare and Grosskopf (2000) (shown below in program 4.15). In this paper the authors propose the division of total output into a part that is final and a part that is kept in the system to be used in subsequent time periods. In this specification, we use a radial output expansion factor. Note that Fare and Grosskopf (2000) only propose a technology for dynamic models and do not discuss an efficiency measure. The use of the output radial expansion with this technology set has been proposed by Kao (2013), and we decided to follow this strategy to make the discussion more clear. If we were to evaluate the efficiency of the input-output combination $(x_{npo}, y_{mpo}, z_{lpo})$ (where o is indexing the DMU under evaluation), the program would be:

$$\begin{aligned}
&\max_{\lambda_{pj},\theta} \theta \\
&st \quad \sum_j \lambda_{pj} x_{npj} \leq x_{npo}\,, \ \forall n, p \\
&\qquad \sum_j \lambda_{pj} y_{mpj} \geq \theta y_{mpo}\,, \ \forall m, p \\
&\qquad \sum_j \lambda_{pj} z_{l(p-1)j} \leq z_{l(p-1)o}\,, \ \forall l, p \\
&\qquad \sum_j \lambda_{pj} z_{lpj} \geq z_{lpo}\,, \ \forall l, p
\end{aligned} \tag{4.15}$$

In this program, the process index p can be interpreted as time and the intermediate factor (z_l) enters the network at the beginning node 0 and exits at node P. The index p can stand for time or for process, since dynamic models are identical to series models. Therefore, the flow of the intermediate in this network is sequential, flowing from $p = 1$ to $p = 2$, $p = 2$ to $p = 3$ and so on, until reaching node P where it exits as an output. The last two constraints on the intermediates allow for production feasibility by making sure that the activation level at node p is not using more intermediate input $(z_{l(p-1)o})$ than is available and is producing at least the observed amount of intermediate output (z_{lpo}). The reader can convince herself that by appropriately expanding Koopmans' matrices to make all inputs and outputs process specific, program (4.15) becomes a special case of the KKJ model.

In program (4.15), output is maximized by keeping the level of the inputs at the observed level without allowing for reallocation of resources across the different nodes of the system. In Bogetoft et al. (2009) or Färe et al. (2018) the authors call model (4.15) the static model, where intermediates are treated as normal inputs and outputs. When they are considered as decision variables the dynamic nature of the system emerges, given that optimal allocation is determined. Kao (2013) proposes an alternative model in which the system is optimized as a whole, given constraints on the overall quantities of inputs. This means that reallocation of resources across the different nodes is possible and the program becomes:

$$
\begin{aligned}
&\max_{\lambda_{pj},\theta} \theta \\
&st \quad \sum_p \sum_j \lambda_{pj} x_{npj} \le \sum_p x_{npo}\,, \ \ \forall n \\
&\qquad \sum_p \sum_j \lambda_{pj} y_{mpj} \ge \theta \sum_p y_{mpo}\,, \ \ \forall m \\
&\qquad \sum_p \sum_j \lambda_{pj} \left(z_{lpj} - z_{l(p-1)j}\right) \ge z_{lPo} - z_{l0o}\,, \ \ \forall l
\end{aligned}
\tag{4.16}
$$

We should notice here that Kao (2013) is actually calling θ the "system efficiency". This means that the author is actually recognizing that these are "system efficiency" models. This specification can clearly be embedded into the KKJ model by noting that the intermediate material is really nothing more than a resource stock that can be depleted in time. Note that in this specification there is no constraint stating that the intermediate input that enters node p $(\sum_j \lambda_{pj} z_{(p-1)j})$ must be lower than the output exiting node $p-1$ $(\sum_j \lambda_{(p-1)j} z_{(p-1)j})$. This is fine, as long as the overall underlying stock variable is big enough to compensate any shortages in a given time period (see Kao's paper for an empirical example). Nevertheless, the most important characteristic of this model is that there is a measure of efficiency from the system perspective. In all respects, this model is a special case of the KKJ model, where the intermediate materials are interpreted as a depleting stock. In fact, as we shall see, if we omit the constraint on the depleting resources, this program is equivalent to the industry efficiency model (see model 4.19 in the next section). Such a model is also proposed in Kao (2012) for parallel production systems (which resemble in all respects industry models). Note that models (4.15) and (4.16) evaluate efficiency relative to different technologies. While in (4.15), technology is process/time dependent—i.e., there is a technology considered individually for each process or time period, in (4.16) a process meta-technology is employed, where the objective function does not yield process/time-specific efficiencies, but the efficiency of the system as a whole, like in the KKJ model. In order to recover process efficiency scores from the system efficiency score, Kao (2013) proposes to use the multiplier form and the associated optimal multipliers for deriving process/time period-specific efficiencies. This results in some problems of the approach, one of which related to the fact that multipliers are not unique and another being the inconsistency between targets obtained from the envelopment model and the efficiency scores obtained from the multiplier model.

Series Network Models

Series models are a special case of dynamic models where different processes within the same firm are connected through intermediate factors and inputs and outputs consumed at different stages may be different. For some reason, the literature has given particular emphasis to two-stage series models where the main focus has been the analysis of the aggregation of process efficiency scores. The general model presented in Kao (2014) for handling multi-stage series models is similar to model (4.16) for dynamic models. In the Kao (2014) model (4.15) is also proposed as an alternative method for solving series network models where "the technologies of all processes are allowed to be different" (p. 2). Note that the differences between these two models do not relate only to different technologies, but also different treatment of intermediates. Tone and Tsutsui (2009) noted that one can have two assumptions on the intermediate variables. The fixed link assumption is stating constraints on the intermediates as in program (4.15) by constraining these variables to their observed level. In the free link approach, they assume that the intermediate variables can be freely chosen (given some feasibility constraints as in the KKJ model). The model proposed by Tone and Tsutsui (2009) is similar to model (4.15), except that they use a slack-based efficiency measure and use equality variables for the intermediates, where under the free link intermediate factor targets from the succeeding stage are set equal to targets for the preceding stage, and under the fixed link the targets for intermediates are set equal to observed values. Fukuyama and Mirdehghan (2012) analyzed this model and concluded that the approach of Tone and Tsutsui (2009) did not account for inefficiencies from intermediate factors.

Kao and Hwang (2010) were among the first to propose two-stage network models. Under this special case of series network, the structures of models in the literature are very similar to those of models (4.15) and (4.16). The difference is that in this case where inputs and outputs are different across stages the use of the meta-technology is not possible and most models resemble model (4.15), with differences mainly in the treatment of intermediates. For example, Lim and Zhu (2016) or Chen et al. (2013) propose two-stage models where intermediates are decision variables, similarly to what is proposed in Nemoto and Goto (2003) (and similar to the free link approach).

The literature on the two-stage models is mainly concerned with decomposing the overall efficiency of the firm into stage 1 and stage 2 efficiencies. Kao (2013) provides some decomposition between process efficiencies and firm efficiency for the case of series systems. Various types of decompositions exist in the literature with the additive and multiplicative ones being the prevalent. Despotis et al. (2016) and Sotiros et al. (2019) point out existing problems and inconsistencies with the original decomposition such as the fact that the maximum firm efficiency score can be obtained from process efficiency scores that are not on the Pareto-frontier, and that could therefore be improved. They propose alternative approaches to solve the problem based on multi-objective linear programming. Li et al. (2018) also analyze two-stage models and provide alternative models for defining which of the processes is the leader or the follower.

Parallel Network Models

Parallel network models can be seen as a set of processes within a DMU that may share some resources. The main feature that distinguishes parallel networks is that there is no flow of intermediate materials between the processes. As a result, parallel networks can be represented using model (4.16), where the constraints regarding intermediates are deleted. Kao (2012) proposed this definition of parallel network models and studied them using the following multiplier form:

$$\begin{array}{ll} \max\limits_{u_m, v_n} & \sum_m v_n \sum_p x_{npo} \\ st & \sum_m u_m y_{mpj} - \sum_n v_n x_{npj} \leq 0\,, \ \ \forall p, j \\ & \sum_n u_m \sum_p y_{mpo} = 1 \end{array} \tag{4.17}$$

This is the dual of program (4.16), with the caveat that intermediate constraints have been omitted since we are dealing with a parallel network. The only difference with the program presented in Kao (2012) is that we are using output orientation instead of input orientation: this choice simplifies the discussion and makes the connection to the previous sections more transparent. In program (4.17), u_m is the weight assigned to output m and v_n is the input weight assigned to input n—weights are considered the same across subunits (i.e., the implicit value attributed to each input and output should be the same in each sub-unit). Note that the original model of Kao (2012) has more constraints, but some are redundant. As a result, we simplified it by excluding redundant constraints and ignoring slacks. This results in model (4.17).

According to Kao (2012), model (4.17) results in efficiency scores for each DMU_o (E_o^*). The efficiency of sub-unit p in DMU j (e_{pj}) is determined using the optimal weights of model (4.17) (indexed with a $*$ that means they are the optimal values from program [4.17]):

$$e_{pj} = \frac{\sum_m u_m^* y_{mpj}}{\sum_n v_n^* x_{npj}} \tag{4.18}$$

The computation of subunit efficiencies in this way allows the DMU efficiency to be decomposed into the efficiency of the subunits, using appropriate weights. Being the dual of a model that is nested in the KKJ model implies that Kao (2012) is basically proposing to use the shadow prices associated with the KKJ model to assess the efficiency of the individual production units. This is in line with the intuitions provided by

Kantorovich in his book and the use of shadow pricing as a decentralization mechanism to reach efficiency in Koopmans. If one were to allow production units to trade at the Kao (2012) prices, this would implement the central plan in a decentralized manner (as suggested by Kantorovich and Koopmans).

The use of a meta-technology in the parallel network model has some implicit assumptions, not always explicitly discussed. In particular the assumption that all inputs are perfectly allocatable. Under CRS, this assumption implies that the meta-frontier will be constituted by the most productive process, which implies at the optimal solution that inefficient processes are advised to closure (see also Pachkova, 2009). This provides inconsistencies between the multiplier and envelopment formulations since in the multiplier model all processes will have an efficiency score, where in the envelopment model targets for some processes will be zero. Most of these problems derive from misconceptions regarding what model (4.17) is supposed to measure. It is a firm model, assessing the average unit and assuming that complete reallocation of resources is possible (e.g., closure of some processes to replace them by the most efficient ones). In the disaggregation of the system efficiency proposed in (4.18) the reallocation of resources within firms is disregarded and the whole is considered the sum of the parts. In Peyrache and Silva (2019), these issues are discussed and the authors maintain that firm efficiencies are not simply the sum (or product) of processes efficiencies but include a reallocation component that is mostly disregarded in the literature.

Note that an alternative to solving parallel models would be to use model (4.15) without the intermediate constraints. This solution is not without problems too. In fact, a single expansion factor is used across processes in this model, implying that the solution equals the maximum of process efficiencies as assessed independently (which may be an inadequate aggregate measure for the firm).

As we will see in the next sections, the application of model (4.15) to parallel network models is closely linked with the literature on output-specific inputs and the application of model (4.16) is closely linked with the literature on industry models.

Multi-Level or Hierarchical Models

The term multi-level model has been used by Castelli et al. (2010) and Castelli and Pesenti (2014) to mean the assessment of production units at different levels of decision making. Another term has been used by Cook et al. (1998) and Kao (2014) to mean the same thing: hierarchical models. In this sort of models, firms are grouped into hierarchies, where for example different factories belong to the same plant, and different plants belong to the same company, and different companies belong to the same industry. In this type of models, the problem is that of aggregating the efficiency of factories to obtain the efficiency of plants and then aggregate the efficiency of plants to obtain the efficiency of companies, given that there may be inputs and outputs that are level specific. So these models include the industry models or centralized allocation model, where the problem is exactly the same: to aggregate firm efficiency to get the industry global or structural efficiency.

Multi-level structured data (data that are observed at a system level and cannot be disaggregated in lower levels) may arise in many settings. For example, in education grades are available at the student level, but the number of teachers is available at the school level. Multi-level data is in fact related to group frontiers and meta-frontiers (see, e.g., O'Donnell et al., 2008) where individual firms are usually grouped according to a higher-level characteristic (students may be grouped in private schools and public schools, firms may be grouped according to location or district, etc.). In this type of models, higher-level variables enter the analysis in the constitution of the homogeneous groups, but not as inputs or outputs of the higher-level production process.

In Cook et al. (1998), the authors consider different levels for the variables and solve multiplier models with different multiplier factors associated with each level. When solving the higher-level model, they include constraints for that level and also for the other levels, such that the optimal solution of multipliers for the higher level can also be applicable at lower levels. They assume that the higher-level variables are not allocatable. Cook and Green (2005), or Cook and Green (2004), assumed that these higher-level variables are allocatable, and the model resembles the one presented in Beasley (1995) (which we will refer to later on under [iii]). Castelli et al. (2004) also proposed models for hierarchical structures, but rather than being multi-level models, these are models with a series structure within a parallel structure.

Industry Models

The literature on industry models tries to aggregate the efficiency of each constituent firm to form the efficiency of the industry (or structural efficiency). It started as early as the work of Farrell (1957) and has been discussed also by Forsund and Hjalmarsson (1979) who advocate the use of the average firm for measuring structural efficiency. Ylvinger (2000) advocates that the average unit assessment is not equivalent to the efficiency of the industry, and Li and Cheng (2007), showed that the weighted average of firm efficiencies and the efficiency of the average unit are equivalent concepts under an identical convex individual technology set, and that differences between the two are related to allocative efficiency. Karagiannis (2015) explored in more depth the relationship between the efficiency of the average unit and structural efficiency. The authors conclude that the two concepts of efficiency will coincide only if size is uncorrelated with efficiency and if there are no reallocation inefficiencies. The efficiency of the average DMU has been explored by several authors under the denomination of "Industry models" (e.g., Lozano & Villa 2004; Peyrache & Zago (2016; Peyrache (2013, 2015), where allocation issues between firms in the industry are usually at the center of the discussion. Kuosmanen et al. (2006) also proposed similar models for analyzing the industry cost efficiency and named them top-down approaches. Note that industry models are also related to input-output tables which can be seen as industry models where the industry is an economy composed of various sectors of activity (see Prieto & Zofio, 2007).

The centralized resource allocation model discussed by Lozano and Villa (2004) somehow epitomizes the core of both the industry models and the multi-level models. We therefore discuss it a little more in depth. The model is presented in Lozano and Villa (2004) in input orientation under VRS. If we were, for sake of comparison, switch to output orientation, then the model would be:

$$\begin{array}{ll} \max\limits_{\lambda_{pj},\theta} \theta & \\ st & \sum_p \sum_j \lambda_{pj} x_{nj} \leq \sum_j x_{nj}\ , \ \forall n \\ & \sum_p \sum_j \lambda_{pj} y_{mj} \geq \theta \sum_j y_{mj}\ , \ \forall m \end{array} \tag{4.19}$$

where the P nodes of the system are the firms (which are also used in the definition of the technology). Model (4.19) is equivalent to Kantorovich's problem C, reported in equation (4.3), by noting that we impose the restriction $x_{npj} = x_{nj}, y_{npj} = y_{nj}$ on the data matrices (i.e., all processes uses the same technology) and by setting $g_m = \sum_j y_{mj}$, $\chi_n = \sum_j x_{nj}$ and $q_p = 1$. In other words, the three-dimensional Kantorovich matrix of data is simplified by assuming $x_{npj} = x_{nj}$ and $y_{mpj} = y_{mj}$, with $P = J$. This specification is also equivalent to a model where the efficiency of a virtual DMU with average inputs and outputs is assessed. In fact by dividing all constraints (left and right hand sides) by the number of firms J, one would obtain the average firm interpretation. The assessment of this average unit was first proposed by Forsund and Hjalmarsson (1979) for measuring the structural efficiency of an industry (see also Ylvinger, 2000). The solution of the model under this specification can yield results that are *prima facie* contradictory, since it is possible for an industry to be composed of only technically efficiency units (i.e., when assessed individually they all lie on the frontier) and, at the same time, the industry (composed by these technically efficient units) may be inefficiently organized (see, e.g., Ylvinger, 2000). Indeed, what happens is that when the average unit is used for assessing the industry, reallocation of resources is implicitly considered possible and therefore each firm may individually be performing at its best, but reallocations within the industry could still improve its overall efficiency (i.e., output). This is supposedly one of the reasons for Lozano and Villa (2004) calling their models centralized resource allocation models—since resource allocation between firms is at the heart of such models (see also Mar Molinero et al., 2014). Issues of aggregation and decomposition are also addressed in these models, particularly when they are used to assess industry structural efficiency. For example, Li and NG (1995) show that structural efficiency equals the product of aggregate efficiency and a component of reallocation efficiency, and Karagiannis (2015) decomposed additively structural efficiency into aggregate efficiency (or average efficiency) and a covariance term relating deviation in output shares and technical efficiencies from their averages.

Allocability Models

The last class of models that we want to discuss deals with the explicit definition of different types of inputs and/or outputs. This is a major issue in network models, since once the black box of production is open, one has to state which inputs can be allocated, which ones cannot and which ones are only available at higher level of aggregation. The precursor of these models can be considered the work of Fare (1986) since the author assumes that for some inputs time allocation is not known, and this allocation could be derived. The idea of unknown allocation of otherwise allocatable inputs was used by Beasley (1995). Castelli and Pesenti (2014) call this model the shared flow model. Beasley (1995) assesses the efficiency for two types of university functions (teaching and research) that have specific inputs and outputs but also share some inputs whose allocation is unknown. The author assesses the two functions separately and then considers the determination of the optimal allocation of the shared resource between the two functions (see Ding et al. 2015) for a recent review of this strand of literature). The most important feature of this models is that it implies a (a priori) classification of inputs (some are allocatable or shared between functions/processes and others are not). Following the same idea, in the output-specific input literature, different technologies are associated with different sets of inputs and outputs, and one cannot assume that all inputs are used in the production of all outputs. Cherchye et al. (2013) and Cherchye et al. (2017) propose models that can handle process-specific and shared inputs (or "joint inputs" as they named them). These models assume that joint inputs are simultaneously used by all processes and cannot be distributed (or allocated) to the different processes. Recently, Podinovski et al. (2018) propose a Multiple Hybrid Returns to Scale (MHRS) technology where it is assumed that shared inputs are allocated to different processes (in spite of the allocation not being observed).

Shared flow models imply the existence of shared allocatable resources, but the allocation is not observed (or there is no *a priori* information on the allocation). These models yield an efficiency score that is different from what would be obtained if one assumed that the shared resource was fully available to each process. But this difference only exists because *prior* information on allocation is provided through the form of weight constraints. Therefore, these models seem to classify shared resources into one category that is somewhere in the middle between "Full information on resource allocation is observed" and "No information on resource allocation is observed", which should be the category "Partial information on resource allocation is known/desired". The literature has also been very confusing on this matter as no such classification exists so far.

Summing Up

In general terms, we may characterize existing system models according to the technology employed, the treatment of intermediates and the type of efficiency measure proposed. For example, industry models and parallel network structures may be defined in relation to a meta-technology (the intersection of processes technologies), or in relation to process-specific technologies. Indeed, the model of Kao (2009) for parallel networks resembles the centralized industry model presented in Lozano and Villa (2004). In this type of model, the assessment is equivalent to "finding common input and output weights that maximize the efficiency of a virtual DMU with average inputs and outputs" (Lozano & Villa, 2004, p. 149). On the contrary, process-specific technologies, as those applied in output-specific input settings, in general yield the efficiency of the DMU as being the same as the maximum efficiency across its processes (and therefore, disregard completely inefficient processes).

Process-specific technologies can be also encountered in series models. The reason is in general obvious—if we have two stages, one consuming inputs and another producing outputs, then the assessment of each stage implies the consideration of process-specific technologies since variables are different in each stage. Interestingly, this does not happen in dynamic models, where in fact the variables repeat in each stage. This is the main reason behind two main ways available in the literature for assessing the efficiency under dynamic models: the Fare and Grosskopf (2000) model and the Kao (2013) model. Most existing models for dynamic network structures use the Fare and Grosskopf (2000) process technologies (or time-specific technologies) like those of Nemoto and Goto (2003) or Tone and Tsutsui (2014), but Kao (2013) models aggregate across time the DMUs inputs and outputs (and therefore use a meta-technology).

Another major distinction that one can find between models in the literature is on the treatment of intermediates. Tone and Tsutsui (2009) provide an interesting classification for intermediates: the free link approach and the fixed link approach. Most of the existing models use one way or another for dealing with intermediates. The main difference between them lies in the consideration of inefficiency sources on the use of intermediates in the overall efficiency of the DMU or not. Fukuyama and Mirdehghan (2012) noted this problem in relation to the Tone and Tsutsui (2009) model that did not include inefficiencies from intermediates and provided a way to fix that. Indeed, the type of efficiency measure considered is probably the major difference between models in the literature (e.g., in series models several papers exist that show different ways of computing overall efficiency and aggregating processes efficiencies as testifies Cook et al. [2010] in their review).

All in all, from our review of this large body of literature, we found that the distinctions between the KKJ model and the various strands of literature described in the previous sections are really minor. Most (if not all) of these differences reside in the definition of the efficiency measure. All of the other issues associated with allocability or not of inputs and outputs are really relegated in the building of appropriate data matrices in the KKJ model.

Topics for Future Research

As we saw in the previous sections, one way of rationalizing the growing body of literature on Network DEA models is to look at it from the perspective of the KKJ model. In this sense, the main problem is shifted from the measurement of firm level efficiency to the measurement of the efficiency of the system as a whole and attributing efficiency to possibly the different levels or hierarchies in the system. By looking at this literature from this perspective, one has also the advantage point of making connections to other methodologies in engineering that deal with allocation of resources. In fact, the KKJ model is useful to determine the level of inefficiency of the system, but the input and output targets set by the model can have multiple solutions. The literature is quite silent on how we choose among these alternative allocations, and ideas from the system thinking may help in selecting appropriate and realistic targets in each particular situation. In the rest of this section, we will look into what we think are the open problems associated with the KKJ model and therefore Network DEA models. As we saw, the field of productivity and efficiency analysis developed in the first 30 years (1939–1972) around the KKJ model; it then turned its attention to firm level efficiency estimation for another 30 years (1977–2001); although some papers dealt with resource allocation during this time, it is really only in the last 20 years that the field has been re-discovering the KKJ model and started progressing to solve some inherent problems associated with that type of modeling. In what follows, we are going to present an overview on the main problems associated with the KKJ model.

Efficiency Measurement vs Structure of the Network

It is possible to design the data matrices of the KKJ model in order to accommodate a great deal of network structures. In principle, one should separate the building of the technology reference set, from the measure of efficiency that can be used to measure the inefficiency of the system. Given that the reference set can be represented in a compact way by designing the associated input and output data matrices appropriately, in this section we consider what type of efficiency measure one should use. The literature developing in the last 20 years, as one would expect, has used both radial and slack-based measures of efficiency. From the point of view of our argument, the choice between these two classes of measures does not present any additional challenges compared to a simple and standard DEA model. Russell and Schworm (2009, 2011) have shown that from an axiomatic point of view the two measures of efficiency can be rationalized by looking at the axioms that they satisfy. In particular, radial measures will satisfy continuity, while slack-based measures will satisfy indication (Pareto efficiency). Depending on the particular application, one may choose one measure or the other, but the fact that we are dealing with a network structure is not really adding any additional arguments in favor of one or the other. The only additional argument one has to keep in mind is that hierarchical network models have decision making happening at various levels. Therefore, there is an issue of simplicity of aggregation of the measure of efficiency. In this sense, using a measure of efficiency which is simpler to aggregate will provide an easier way of assessing the efficiency of the system and its components.

Unobserved Allocations

In the KKJ model and in general in the recent Network DEA models, it is assumed that the allocation of the various inputs and outputs is observed. For example, if a firm is composed of P production plants, one observes in the dataset the allocation of each input and the production of each output at each node of the system. What happens if these allocations are not observed or only partially observed? Suppose that the allocation of raw materials to each different node p is observed, but the allocation of labor is not observed. In other words, suppose that we have a case where we know that a given input (labor for example) is allocatable, but we do not observe its allocation. Although this is likely to be a very common case in practice, the literature is quite silent on this point. Once could treat the input as a public good (a joint input), but this is clearly introducing a bias in the measurement of efficiency. Podinovski et al. (2018) propose a solution to this problem for the case of CRS technologies. The reader should refer to this very important contribution to gain a better perspective of the modeling strategy. For our purposes, it suffices to say that the Podinovski et al. (2018) model provides a technology reference set that is contained in the one that one would obtain if the allocations were observed. This has the great advantage of providing a conservative estimate of the inefficiency of the system. Extensions to VRS and other scale characterizations are yet to be made. In the absence of a model that extends the ideas of Podinovski et al. (2018) to the VRS case, one could use a suggestion of Farrell (1957). This consists of dividing up the dataset in clusters of observations that have the same "size" and then apply the Podinovski et al. (2018) model to these classes. Although this is less satisfactory than an extension of Podinovski et al. (2018) to the VRS case, it is really the only viable option to deal with unobserved allocations, unless one is willing to interpret the input as a public good (joint input).

In a very recent paper, Gong and Sickles (2021) adapt the stochastic frontier model to the case of a simple parallel network (they use a different wording). This paper is important in itself just because is the first attempt to propose a network model in the stochastic frontier tradition. But for our discussion it is also important because it is dealing with unobserved allocations of allocatable inputs. In particular, the authors make use of input price information to make inferences about the possible allocation of inputs across the different processes. Although the study assumes that price information is available, this is a first attempt at dealing with the problem in a stochastic frontier framework.

Costly Reallocation

In the KKJ model and subsequent work on Network DEA, it is implicitly assumed that either reallocation of inputs is not possible (i.e., inputs are process specific), or reallocation of inputs can happen at no cost. What if the reallocation is costly, but not prohibitively so? To the best of our knowledge, there is only one paper dealing with costly reallocations (Pachkova, 2009). This is likely to be a very important problem in practice, since reallocation of resources is likely to happen at some cost. In particular, one can look at inputs that are specific to a particular process as a resource that is allocatable but the cost of reallocation is either very high or prohibitive. For example, if we think of beds in a hospital, they are likely to be allocatable at negligible cost (i.e., the cost of transporting them from one department to another or one hospital to another). On the contrary, doctors, given their specialization, are unlikely to be allocatable at no cost. Even if one could retrain a cardiologist to become a radiologist, this is likely to take a lot of time, money and effort. Therefore, in the short run, at least the number of doctors in a hospital represents an input that is prohibitively costly to reallocate. In general terms, if information on the cost of reallocation is available, one should be able to introduce it into the KKJ model in order to take it into account. In this way, the model becomes a hybrid transportation-production model, where optimality is reached taking into account the actual possibilities and costs of reallocation of resources.

Connection Between Network Analysis and the Black Box Analysis

What happens if we run the analysis at the black box level rather than the network level? One formal way of stating this is the following. Call T_p the production possibilities set of process p and each process is allocated input x_p to produce output y_p. The total for the firm is $X = \sum_p x_p$ and $Y = \sum_p y_p$. The firm production possibilities set is given by all the possible allocations of the inputs across the different P processes:

$$T = \left\{ \left(\sum_p x_p, \sum_p y_p \right) : \left(x_p, y_p \right) \in T_p \right\} \tag{4.20}$$

Suppose now that we run the analysis at the firm level and we build the production possibilities set using the total inputs and outputs of the firm. Call this set T_F. What is the relationship between T and T_F? In other words, if we know that the firm is composed of different departments (cardiology, radiology, etc.) but we run the analysis at the firm level ignoring the allocations to the various departments, can we still obtain meaningful efficiency scores? Is it possible to make general statements about their relationship? For example is the black box technology always underestimating efficiency? In general, we think the answer is no, and convexity plays a big role in addressing this issue. Is it possible to have general results? We found only one theoretical paper by Buccola and Fare (2008) dealing with this issue. This is actually an important area of research, since it is connected with the simplification of the analysis and it makes an explicit connection between the black box technology and the underlying process-specific technologies. In general, one may want to build T_F in such a way that it is contained in T. If so, the estimation of efficiency at the firm level using the black box technology is higher than the one estimated using T and this means that the KKJ model helps to increase discrimination power. In general, conditions for this to happen will involve some restrictive assumptions that we still don't know.

Network Stochastic Frontiers

This is possibly the biggest missing point in the literature. With the exception of Gong and Sickles (2021), we could not find a single stochastic frontier paper that is dealing with some form of network structure. Stochastic frontier analysis applied to network production structures can bring about many benefits. Although the standard narrative is to say that the difference between SFA and DEA is coming from the noise component, it is important to stress that SFA allows the introduction of functional forms. If the dataset has a small number of observations, then it makes sense to parameterize the production frontier function and assume that it has some known parametric form. In general, SFA analysis may provide an advantage in this sense. One may use SFA as a noise-canceling device and once estimation is done, use the estimated coefficients to determine the optimal allocation of resources. As long as the functional form is convex, the KKJ would become a convex program rather than a linear program. Convex programming made some strong progress in computational terms. If one wants to stick with linear programming, then it is possible to follow the suggestion of Koopmans (1951) of approximating the known functional form with a piece-wise function. In fact, one could go a step further and estimate directly a spline function in a SFA framework and use it to retain a linear program specification for the KKJ model.

Micro-foundation of the Aggregate Production Function

Johansen (1972) had a chief interest in the micro-foundation of the macro- or aggregate production function. The KKJ model was interpreted by Johansen as a tool to describe the aggregate production possibilities set starting with observations at the micro-level or, in other words, from observation of the firm level input-output combinations. In this respect, the KKJ model is a nonparametric way of determining the aggregate production function. By making specific assumptions on the statistical distribution of the inputs and outputs across firms, one can infer specific functional forms for the aggregate production function. Johansen discusses a number of them. If we were to take this approach to its logical consequences, then one should start with the estimation of the distribution functions of the inputs and outputs and once these distributions are known determine the aggregate production function. This would open up the way to the use of flexible ways of estimating multivariate distribution functions such as copulas. Work in this space is very much limited, to the best of our knowledge, to the proposals of Johansen. Given the progress that has been made in the last 50 years in terms of estimation of multivariate distribution functions, it is quite clear that this is now a viable and potentially very fruitful avenue of research that is underexplored. The intuition of Johansen can be given more explicit content and it would be possible to specify a number of alternative ways of extending this idea to the more general setting of the KKJ model.

Epilogo

The previous pages provide a number of important unexplored topics that are relevant to the modern researcher in efficiency and productivity analysis, especially if she is willing to focus on problems associated with central planning and regulation of markets. We also provided a brief history of this field of study, and hopefully, we have provided evidence that many of the NDEA models developed in the last 10 years or so are just special cases of the KKJ model that can be dealt with by adjusting in a proper way the data matrices as presented in Koopmans and Kantorovich. As a result of separate developments, each of the above strands of literature tends to look at the same problem from different perspectives, like in the Indian elephant parable where each blind man guessed a different object depending on the body part of the elephant they were sensing (see Fig. 4.1)

Given the current status of this field of research, the themes proposed in the last section to progress forward this field are unlikely to be explored at the same pace at which the NDEA literature has been growing in the last 10 years. This may be due to a number of factors, many of them having to do with the way research is structured today. A question one

Fig. 4.1 Parable of the blind men and an elephant originated in India

should really ask is why the sub-fields described in section Summing Up have been growing into almost completely separate streams of literature, even if they all are under the umbrella of system efficiency analysis (and mostly just variations of the KKJ model). In this last subsection, we should speculate on how the field arrived at such a state of affairs.

Clearly, the working environment of the modern researcher is very different from the one in which academics used to work in the past. The pressure to publish papers has become bigger and bigger. Universities value research output based on quantity rather than quality, in most cases. This means that researchers have a strong incentive to engage in salami slicing (the practice of taking a single piece of research and fragment it into smaller pieces that can be published). This prompted Wikipedia to have a page describing what this is (search on Wikipedia for "least publishable unit"). The "least publishable unit" has become definitely smaller in time. The interested reader can make a quick Google Scholar search with the keyword "publish or perish" to see that there are already a number of papers concerned with the distortions that this system is producing.

Universities require academics to be "leader" in their own field of research. This means that academics have a strong incentive and a tendency to create sub-fields and over-represent their contribution within these sub-fields. In particular, many of these sub-fields are not even so different from each other, at least for what we saw in the previous few sections. This state of things is creating a dangerous mentality, and we are breeding an entire generation of researcher that hyper-specialize, by teaching them how to best market their research in order for it to look original, so they can be "leader" in their respective fields of research. The quest for truth and knowledge has been replaced by the quest for publication at any cost. The collegiality and intellectual honesty of the scientific international community have been replaced by a grim citation count. This basically transformed international conferences from places where academics share and progress knowledge, into places where researchers put forward aggressive marketing campaigns (sometimes on the edge of bullying and harassment) to increase their citation count, h-index and impact factor. Journals have followed this trend, transforming editorial boards into lobbies that look after the "insiders", instead of having their more traditional function of recognizing original and relevant work irrespective of where it is coming from. The ingenuity and fascination of true knowledge that drives many people into the search for academic jobs (and is so much needed for the advancement of truth and knowledge) are quickly replaced by a more mundane need to be competitive on the market for academic jobs. Instead of leaving small details associated with the development of models and results out of the papers, we create entire new papers out of these details. It is quite amusing that by reading Kantorovich work, many small details and intuition were left to "the production engineers" (this resembles the traditional role of the teacher that is leaving some details to be sorted out by the student as homework). Sorting out such details would of course imply that the "production engineers" (to stick with Kantorovich) have a good education in the first place that allows them to do so. Out of these details, we now build entire journals that are trying to "fill the gaps" in the literature. Roger Koenker, notably one of the most creative and prominent econometrician and statistician of our time (and the proponent of quantile regression; another field to which productivity analysis should have closer connections...), has suggested that we should all be part of the "Society for the Preservation of Gaps in the Literature" (the interested reader can visit: https://www.econ.uiuc.edu/roger/gaps.html). To use his words:

> Gaps in the literature constitute the essential breathing spaces of academic life. The research and publication process poses an increasing threat to the well being of disciplines by gradually filling these gaps with meritless

> interpolation of existing results. The Society for the Preservation of Gaps in the Literature is dedicated to the preservation of the "intellectual green space" afforded by these gaps.
>
> Rather than filling gaps in the literature one of the great accomplishments of serious research is to create gaps in the literature by debunking the nonsense of the past. Nowhere is this objective better formulated than in the introduction to the bibliography of Keynes' (1921) Treatise on Probability:
>
> "I have not read all these books myself, but I have read more of them than it would be good for any one to read again. There are here enumerated many dead treatises and ghostly memoirs. The list is too long, and I have not always successfully resisted the impulse to add to it in the spirit of a collector. There are not above a hundred of these which it would be worth while to preserve,–if only it were securely ascertained which these hundred are. At present a bibliographer takes pride in numerous entries; but he would be a more useful fellow, and the labours of research would be lightened, if he could practise deletion and bring into existence an accredited Index Expurgatorius. But this can only be accomplished by the slow mills of the collective judgment of the learned; and I have already indicated my own favorite authors in copious footnotes to the main body of the text.

There are no better words to describe the state of the literature on the system perspective in efficiency and productivity analysis (maybe to describe the state of the literature in general?). We definitely did not read all papers in NDEA and we have no intention to do so in the future, given that the ones we found are only minor incremental progresses to the KKJ model. In fact, it is hard enough to acknowledge that some of the models proposed by one of the authors of this chapter (Peyrache, 2013, 2015) are so close to the KKJ model to make one wonder if they were to be published in the first place or if they should have been left as homework exercises. We are starting to think that we have ourselves destroyed another gap in the literature and made our academic life less green by adding noise to noise (Peyrache & Silva, 2019).

How is it possible that the literature has grown so fragmented, by producing such an exponential growth in the number of published papers that basically deal with the same underlying problem? If every single author were to walk in the same conference room and read their paper, everyone would be reading the same material in a different "language", creating a lot of chatter. We thought that we may call this effect "publication chatter", like the chatter in the conference room. But this time we were smart enough to look out for a paper instead of re-inventing the wheel. We were surprised to find at least 4 papers on this topic (maybe these papers are also filling a gap in our knowledge?). Kozlov and Hurlbert (2006), in the Journal of Fundamental Biology, pushed the idea that we should learn from mistakes of the past; otherwise, we are going to reproduce the same mistakes in the current literature (we could not have said this better!). They cite the 1984 Dean of the Graduate School, Yale University:

> Nowhere in all of scholarship has the book or shorter contribution (the 'paper') become more thoroughly debased than in science ... the principal remedy is for everyone to write fewer and more significant works ... It seems to be a deeply held, quasi-philosophical position among contemporary scientists that publication, and lots of it, is an inalienable right ... it is no longer an honor to get a paper published ... publication of any and all results has become the norm ... the publication process has largely ceased to act as a quality control mechanism ... It is terribly important for students to appreciate the older literature in their field ... For scientists there is a danger that the vast tide of chatter in the current literature may isolate us from our intellectual underpinnings.

Given that researchers themselves don't have incentives to limit the number of published papers, can we still hope for this to be accomplished by the refereeing process? Is this process really conducive to eliminate papers that only marginally contribute to the literature and really incentivize innovation? Lloyd (1985), in The Florida Entomologist, states:

> Read on: "We share the opinion of Hall (1979), Stumpf (1980), and others that anonymous peer reviews may be more costly than beneficial. A system that could allow a reviewer to say unreasonable, insulting, irrelevant, and misinformed things about you and your work without being accountable hardly seems equitable. To some degree the reviewer is indeed accountable- to the editor-but the potential for abuse is still too great to be ignored" (Peters and Ceci 1982); Rules based on "empirical research," for manuscript acceptance are as follows: "Authors should: (1) not pick an important problem, (2) not challenge existing beliefs, (3) not obtain surprising results, (4) not use simple methods, (5) not provide full

> disclosure, and (6) not write clearly" (Armstrong 1982; see also Harlow 1962).
>
> To sum up, our peer/referee system, the piers of our academic sand castle, can sometimes amount to nothing more than an adversarial confrontation where the defendant is presumed guilty, has no counsel or friend in court by arrangement, cannot face his accusers, and there are no qualifications for judges. At other times, it can be the reverse, and a conspiracy of peers in a field to promote the field (and one another), or a network of master(s) and disciples. Shouldn't we find out how bad it really is and try to fix it, and try to anticipate what will happen next to pervert it?

Thomson (1984) writes on the American Scientist:

> Evidently our way of coping with the flow of minor publications is to ignore them, thereby making them even more trivial. All this work therefore represents the most senseless waste, especially when the occasional gem by an unknown author gets lost in the crowd. In short, nowhere in all of scholarship has the book or shorter contribution (the "paper") become more thoroughly devased than in science (although apparently other fields are doing their best to catch up).
>
> These are harsh words, and logically it will behooves any author to add another paper to the list in order to make the point, when the principal remedy is for everyone to write fewer and more significant works (physician, help thyself). But "less is more" may be hard to attain in this area. Publish or perish is deeply embedded in the subculture of science (and God forbid that we should have to find some more valid criterion in order to judge promotions).

It is somehow sad to see that many good researchers in efficiency and productivity analysis are so deeply entrenched with playing a game that is holding the field from progressing at the pace it should. While closing with this pessimistic note, we also notice that a new generation of researchers in productivity analysis is coming to the scene. With the old guard retiring from editorial boards, this will make it harder to publish, but maybe this will re-orient the research effort of the latest generation of researcher in productivity and efficiency analysis toward a more fruitful and useful path. We really hope so. Even if anecdotal evidence suggests the opposite.

References

Agrell, P. J.K., & Hatami-Marbini, A. (2013). Frontier-based performance analysis models for supply chain management: State of the art and research directions. *Computers and Industrial Engineering, 66*, 567–583.

Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production models. *Journal of Econometrics, 6*, 21–37.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*, 1078–1092.

Beasley, J. (1995). Determining teaching and research efficiencies. *Journal of the Operational Research Society, 46*, 441–452.

Bogetoft, P., Färe, R., Grosskopf, S., Hayes, K., & Taylor, L. (2009). Dynamic network DEA: An illustration. *Journal of the Operations Research Society of Japan, 52*(2), 147–162.

Buccola, S., & Fare, R. (2008). Reaggregation and firm-level inference in multiplant technologies. *Journal of Economics, 95*(3), 255–270.

Castelli, L., & Pesenti, R. (2014). Network, shared flow and multi-level DEA models: A critical review. In W. Cook & J. Zhu (Eds.), *Data Envelopment Analysis, International Series in Operations Research and Management Science 208* (pp. 329–376). New York: Springer.

Castelli, L., Pesenti, R., and Ukovich, W. (2004). DEA like models for the efficiency evaluation of hierarchically structured units. *European Journal of Operational Research*, 154:465?476.

Castelli, L., Pesenti, R., and Ukovich, W. (2010). A classification of DEA models when the internal structure of the decision making units is considered. *Annals of Operations Research*, 173:207?235.

Charnes, A., & Cooper, W. W. (1962). On some works of kantorovich, koopmans and others. *Management Science, 8*(3), 246–263.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research, 2*, 429–444.

Chen, Y., Cook, W., Kao, C., and Zhu, J. (2013). Network dea pitfalls: Divisional efficiency and frontier projection under general network structures. *European Journal of Operational Research*, 226(3):507?515.

Cherchye, L., Rock, B. D., Dierynck, B., Roodhooft, F., & Sabbe, J. (2013). Opening the black box of efficiency measurement: Input allocation in multi-output settings. *Operations Research, 61*(5), 1148–1165.

Cherchye, L., Rock, B. D., & Hennebel, V. (2017). Coordination efficiency in multi-output settings: a dea approach. *Annals of Operations Research, 250*, 205–233.

Cook, W., Chai, D., Doyle, J., & Green, R. (1998). Hierarchies and groups in DEA. *Journal of Productivity Analysis, 10*, 177–198.

Cook, W., & Green, R. (2004). Multicomponent efficiency measurement and core business identification in multiplant firms: A DEA model. *European Journal of Operational Research, 157*, 540–551.

Cook, W., & Green, R. (2005). Evaluating power plant efficiency: a hierarchical model. *Computers and Operations Research, 32*, 813–823.

Cook, W., Liang, L., & Zhu, J. (2010). Measuring performance of two-stage network structures by dea: a review and future perspective. *Omega, 38*(6), 423–430.

Despotis, D. K., Koronakos, G., & Sotiros, D. (2016). Composition versus decomposition in two-stage network dea: a reverse approach. *Journal of Productivity Analysis, 45*, 71–87.

Ding, J., Feng, C., Bi, G., Liang, L., & Khan, M. (2015). Cone ratio models with shared resources and nontransparent allocation parameters in network dea. *Journal of Productivity Analysis, 44*, 137–155.

Fallah-Fini, S., Triantis, K., & Johnson, A. L. (2014). Reviewing the literature on non-parametric dynamic efficiency measurement: state-of-the-art. *Journal of Productivity Analysis, 41*(1), 51–67.

Fare, R. (1986). A dynamic non-parametric measure of output efficiency. *Operations Research Letters, 5*(2), 83–85.

Fare, R., & Grosskopf, S. (1996). *Intertemporal production frontiers: with dynamic DEA*. Boston: Kluwer Academic Publishers.

Fare, R., & Grosskopf, S. (2000). Network dea. *Socio-Economic Planning Sciences, 34*, 35–49.

Färe, R., Grosskopf, S., Margaritis, D., & Weber, W. L. (2018). Dynamic efficiency and productivity. In *The Oxford Handbook of Productivity Analysis* (pp. 183–210). Oxford: Oxford University Press.

Fare, R., & Lovell, C. A. K. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory, 19*(1), 150–162.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A, general 120(Part 3):253–281.

Forsund, F. R., & Hjalmarsson, L. (1979). Generalised Farrell measures of efficiency: an application to milk processing in Swedish dairy plants. *The Economic Journal, 89*(June), 294–315.

Fukuyama, H., & Mirdehghan, S. (2012). Identifying the efficiency status in network dea. *European Journal of Operational Research, 220*, 85–92.

Gardner, R. (1990). Lv kantorovich: The price implications of optimal planning. *Journal of Economic Literature, 28*(2), 638–648.

Gong, B. and Sickles, R. C. (2021). Resource allocation in multi-divisional multi-product firms. *Journal of Productivity Analysis*, pages 1–24.

Houthakker, H. S. (1955). The pareto distribution and the cobb-douglas production function in activity analysis. *The Review of Economic Studies, 23*(1), 27–31.

Isbell, J. R., & Marlow, W. H. (1961). On an industrial programming problem of kantorovich. *Management Science, 8*(1), 13–17.

Johansen, L. (1972). Production functions; an integration of micro and macro, short run and long run aspects. Technical report.

Johansen, L. (1976). Lv kantorovich's contribution to economics. *The Scandinavian Journal of Economics, 78*(1), 61–80.

Jones, C. I. (2005). The shape of production functions and the direction of technical change. *The Quarterly Journal of Economics, 120*(2), 517–549.

Kantorovich, L. V. (1939). *Mathematical methods of organizing and planning production*. Leningrad University.

Kantorovich, L. V. (1960). Mathematical methods of organizing and planning production. *Management science, 6*(4), 366–422.

Kantorovich, L. V. (1965). The best use of economic resources. *The best use of economic resources.*

Kao, C. (2009). Efficiency decomposition in network data envelopment analysis: a relational model. *European Journal of Operational Research, 192*(1), 949–962.

Kao, C. (2012). Efficiency decomposition for parallel production systems. *Journal of the Operational Research Society, 63*(1), 64–71.

Kao, C. (2013). Dynamic data envelopment analysis: A relational analysis. *European Journal of Operations Research, 227*(1), 325–330.

Kao, C. (2014). Network DEA analysis: a review. *European Journal of Operational Research, 239*(1), 1–16.

Kao, C., & Hwang, S.-N. (2010). Efficiency measurement for network systems: IT impact on firm performance. *Decision Support Systems, 48*, 437–446.

Karagiannis, G. (2015). On structural and average technical efficiency. *Journal of Productivity Analysis, 43*, 259–267.

Koopmans, T. (1951). Activity analysis of production and allocation.

Koopmans, T. C. (1953). Activity analysis and its applications. *The American Economic Review, 43*(2), 406–414.

Koopmans, T. C. (1960). A note about kantorovich's paper, "mathematical methods of organizing and planning production". *Management Science, 6*(4), 363–365.

Kozlov, M., & Hurlbert, S. (2006). Pseudoreplication, chatter, and the international nature of science: A response to dv tatarnikov. *Journal of Fundamental Biology (Moscow), 67*(2), 145–152.

Kuosmanen, T., Cherchye, L., & Sipilainen, T. (2006). The law of one price in data envelopment analysis: restricting weight flexibility across firms. *European Journal of Operational Research, 170*, 735–757.

Li, H., Chen, C., Cook, W., Zhang, J., & Zhu, J. (2018). Two-stage network dea: who is the leader. *Omega, 74*, 15–19.

Li, S., & Cheng, Y. (2007). Solving the puzzles of structural efficiency. *European Journal of Operational Research, 180*, 713–722.

Li, S.-K., & NG, Y. (1995). Measuring the productive efficiency of a group of firms. *International Advances in Economic Research, 1*, 377–390.

Lim, S. and Zhu, J. (2016). A note on two-stage network dea model: frontier projection and duality. *European Journal of Operational Research*, pages 342–346.

Lloyd, J. E. (1985). On watersheds and peers, publication, pimps and panache (an editorial abstract). *The Florida Entomologist, 68*(1), 134–140.

Lozano, S., & Villa, G. (2004). Centralized resource allocation using data envelopment analysis. *Journal of productivity Analysis, 22*, 143–161.

Mar Molinero, C., Prior, D., Segovia, M., & Portillo, F. (2014). On centralized resource utilization and its reallocation by using dea. *Annals of Operational Research, 221*, 273–283.

Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International economic review*, pages 435–444.

Nemoto, J., & Goto, M. (2003). measurement of dynamic efficiency in production: an application of data envelopment analysis to japanese electric utilities. *Journal of Productivity Analysis, 19*, 191–210.

O'Donnell, C. J., Rao, D. S. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics, 34*, 231–255.

Pachkova, E. V. (2009). Restricted reallocation of resources. *European Journal of Operational Research, 196*, 1049–1057.

Pasinetti, L. L. (1973). The notion of vertical integration in economic analysis. *Metroeconomica, 1*, 1–29.

Peyrache, A. (2013). Industry structural inefficiency and potential gains from mergers and break-ups: a comprehensive approach. *European Journal of Operational Research, 230*(2), 422–430.

Peyrache, A. (2015). Cost constrained industry inefficiency. *European Journal of Operational Research, 247*(3), 996–1002.

Peyrache, A. and Silva, M. (2019). The inefficiency of production systems and its decomposition. working paper, Centre for Efficiency and Productivity Analysis (CEPA) working paper. WP05/2019.

Peyrache, A., & Zago, A. (2016). Large courts, small justice! the inefficiency and the optimal structure of the italian justice sector. *Omega, 64*, 42–56.

Podinovski, V., Olsen, O., & Sarrico, C. (2018). Nonparametric production technologies with multiple component processes. *Operations Research, 66*(1), 282–300.

Prieto, A., & Zofio, J. (2007). Network dea efficiency in input-output models: with an application to oecd countries. *European Journal of Operational Research, 178*, 292–304.

Ray, S. C., & Hu, X. (1997). On the technically efficient organization of an industry: a study of US airlines. *Journal of Productivity Analysis, 8*, 5–8.

Russell, R. R., & Schworm, W. (2009). Axiomatic foundations of efficiency measurement on data-generated technologies. *Journal of Productivity Analysis, 31*(2), 77–86.

Russell, R. R., & Schworm, W. (2011). Properties of inefficiency indexes on< input, output> space. *Journal of Productivity Analysis, 36*(2), 143–156.

Senge, P. M. (1990). *The Fifth Discipline : the Art and Practice of the Learning Organization*. New York: Doubleday/Currency.

Shephard, R. W. (1970). *Theory of Cost and production functions*. Princeton, New Jersey: Princeton University Press.

Sotiros, D., Koronakos, G., & Despotis, D. K. (2019). Dominance at the divisional efficiencies level in network dea: The case of two-stage processes. *Omega, 85*, 144–155.

Thomson, K. S. (1984). Marginalia: The literature of science. *American Scientist, 72*(2), 185–187.

Tone, K., & Tsutsui, M. (2009). Network dea: a slacks based measure approach. *European Journal of Operational Research, 197*, 243–252.

Tone, K., & Tsutsui, M. (2014). Dynamic dea with a network structure: a slacks based measure approach. *Omega, 42*, 124–131.

Ylvinger, S. (2000). Industry performance and structural efficiency measures: Solutions to problems in firm models. *European Journal of Operational Research, 121*, 164–174.

PART II

Income Distributions and Inequality and Insecurity

CHAPTER 5

Modelling Income Distributions with Limited Data

Duangkamon Chotikapanich, William Griffiths, and Gholamreza Hajargasht

Introduction

It is generally recognized that poverty and excessive inequality are socially undesirable. Reducing global poverty so that fewer individuals are deprived of basic needs is a major objective of international agencies. While what constitutes too much inequality is debatable, there is concern about the negative effects of rising inequality on health, crime and other aspects of society. Also, in extreme cases, inequality has led to the overthrow of governments and changes in the international order.

D. Chotikapanich
Monash University, Melbourne, VIC, Australia

W. Griffiths (✉)
University of Melbourne, Melbourne, VIC, Australia
e-mail: wegrif@unimelb.edu.au

G. Hajargasht
University of Queensland, Brisbane, QLD, Australia

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_5

It is important, therefore, be able to monitor changes in inequality and poverty using suitable measurement techniques. For this purpose, modelling and estimation of income distributions and Lorenz curves play an important role. Data available for modelling and estimation can be available in many forms. They may come from taxation data or from a variety of surveys. We focus on modelling and estimation when the data are limited in the sense that they come in grouped form, typically as the proportion of total income allocated to each of a number of groups, ordered according to increasing income, and with a specified proportion of the population within each group. These so-called income and population shares form the basis for estimating inequality through the Lorenz curve.[1] When share data are combined with data on mean incomes, income distributions can also be estimated, and their relationship with Lorenz curves can be exploited.

Data in grouped form are often utilized for large scale projects where inequality and poverty on a regional or global scale are being measured, and where compilation and dissemination of data in a more disaggregated form would be overly resource intensive. An example of such a study is Chotikapanich et al. (2012). Examples of locations where grouped share data are available for researchers are the World Bank's PovcalNet website[2] and that of the World Institute for Development Economic Research.[3]

Our objective is to summarize methods for estimating parametric income distributions using grouped data, to specify the functions needed for estimation for a number of popular parametric forms, and to provide formulae that can be used to compute inequality and poverty measures from the parameters of each of the distributions. In section Concepts, we introduce notation and concepts to be utilized later in the paper. The density, distribution and moment distribution functions that play an important role are introduced, along with poverty and inequality measures whose values can be calculated from estimates of the parameters of income distributions. We also describe the nature of the data that we assume are available. Section Estimation is devoted to estimation. Choice of estimation technique is influenced by whether or not

[1] We will continue to refer to *income* distributions and *income* shares, but recognize that data are often for *expenditure* that can be treated in the same way.

[2] http://iresearch.worldbank.org/PovcalNet/povOnDemand.aspx.

[3] https://www.wider.unu.edu/database/wiid.

group bounds are provided in the available data and on how the data are grouped: fixed group bounds and random population proportions or fixed population proportions and random group bounds. Both minimum distance (MD) and maximum likelihood (ML) estimators are considered, and results are provided for variants of the MD estimators which depend on which "distance" is being minimized. In section Specification of Distributions, Inequality and Poverty Measures, we tabulate the common parametric distributions that have been used to model income distributions; their density, distribution and moment distribution functions, and moments, are provided. Expressions that can be used to calculate inequality measures from the parameters of the different distributions are also tabulated. Expressions for some poverty measures are given in section Concepts; those for the Watts poverty index are tabulated in section Specification of Distributions, Inequality and Poverty Measures. In large projects, involving many countries and many years, MD and ML estimation can be daunting tasks. In section Simple Recipes for Two Distributions, we describe two relatively simple estimators for two specific distributions: the lognormal and the Pareto-lognormal. Some concluding remarks follow in section Concluding Remarks.

Concepts

We assume a population of incomes y, with $y > 0$, can be represented by a probability density function (pdf) $f(y;\theta)$ where θ is a vector of unknown parameters. Our objective is to review several alternative functional forms that have been suggested for $f(y;\theta)$, to describe methods for estimating θ from grouped data, and to provide expressions that can be used to compute estimates of inequality and poverty measures from estimates for θ.

We further assume y has a finite mean $\mu = \int_0^\infty y\, f(y;\theta)\, dy$. Its cumulative distribution function (cdf) will be denoted by

$$\lambda = F(y;\theta) = \int_0^y f(t;\theta)dt \tag{5.1}$$

and its first moment distribution function (fmdf) by

$$\eta = F^{(1)}(y;\theta) = \frac{1}{\mu}\int_0^y t\, f(t;\theta)\, dt \tag{5.2}$$

We will also utilize the second moment distribution function (smdf)

$$\psi = F^{(2)}(y;\theta) = \frac{1}{\mu^{(2)}}\int_0^\infty t^2 f(t;\theta) dt \tag{5.3}$$

where $\mu^{(2)}$ is the second moment $\mu^{(2)} = \int_0^\infty y^2 f(y;\theta)\, dy$. The Lorenz curve, relating the cumulative proportion of income to the cumulative proportion of population, is given by[4]

$$\eta = L(\lambda;\theta) = F^{(1)}\left(F^{-1}(\lambda;\theta);\ \theta\right) \tag{5.4}$$

When modelling begins with the specification of a Lorenz curve, the quantile function $y = F^{-1}(\lambda;\theta)$ can be found from it via differentiation,

$$y = F^{-1}(\lambda;\theta) = \mu \frac{dL(\lambda;\theta)}{d\lambda} \tag{5.5}$$

Inequality Measures

The most commonly cited inequality measure is the Gini coefficient g which is given by twice the area between the Lorenz curve and the line of equality where $\eta = \lambda$. That is,

$$\begin{aligned} g &= 1 - 2\int_0^1 L(\lambda;\theta)\, d\lambda \\ &= -1 + \frac{2}{\mu}\int_0^\infty yF(y;\theta) f(y;\theta) dy \end{aligned} \tag{5.6}$$

[4] See Gastwirth (1971).

Two further inequality measures that we consider are the Theil indices which are special cases of a generalized entropy class of measures. Unlike the Gini coefficient, members of this class have the advantage of being additively decomposable into population subgroups. The general class is given by

$$\mathrm{GE}(v) = \frac{1}{v^2 - v}\left[\int_0^\infty \left(\frac{y}{\mu}\right)^v f(y;\theta)dy - 1\right] \qquad v \neq 0,\ 1 \tag{5.7}$$

The parameter v controls the sensitivity of the index to income differences in different parts of the income distribution; larger positive values imply greater sensitivity to income differences in the upper part of the distribution and more negative values imply greater sensitivity to differences in the lower part of the distribution. The Theil special cases are those for $v \to 0$ and $v \to 1$. They are given by

$$T_0 = \mathrm{GE}(0) = \int_0^\infty \ln\left(\frac{\mu}{y}\right) f(y;\theta)\,dy \tag{5.8}$$

$$T_1 = GE(1) = \int_0^\infty \left(\frac{y}{\mu}\right) \ln\left(\frac{y}{\mu}\right) f(y;\theta)dy \tag{5.9}$$

The last inequality measure that we consider is the Pietra index which is equal to the maximum distance between the Lorenz curve and the equality line $\eta = \lambda$. It can be written as the difference between the cdf and the fmdf, evaluated at μ.

$$P = F(\mu;\theta) - F^{(1)}(\mu;\theta) \tag{5.10}$$

Poverty Measures

Modelling and estimating income distributions are also useful for evaluating poverty. We consider four poverty measures, the headcount ratio HC, the poverty gap PG, the FGT index with the inequality aversion parameter set at 2 and the Watts index, WI. For convenience, we express HC, PG and FGT in terms of distribution and moment distribution functions, and moments, which are tabulated for specific distributions in section Specification of Distributions, Inequality and Poverty Measures. The Watts index requires more work, however; we defer specific parametric expressions for it until section Specification of Distributions, Inequality and Poverty Measures. Given a specific poverty line z, we have

$$H = F(z;\theta) \tag{5.11}$$

$$\mathrm{PG} = \int_0^z \left(\frac{z-y}{z}\right) f(y;\theta)\,dy = F(z;\theta) - \frac{\mu}{z}F^{(1)}(z;\theta) \tag{5.12}$$

$$\begin{aligned}\mathrm{FGT}\,(2) &= \int_0^z \left(\frac{z-y}{z}\right)^2 f(y;\theta)\,dy \\ &= F(z;\theta) - 2\frac{\mu}{z}F^{(1)}(z;\theta) + \frac{\mu^{(2)}}{z^2}F^{(2)}(z;\theta)\end{aligned} \tag{5.13}$$

$$\mathrm{WI} = \int_0^z [\ln(z) - \ln(y)]\, f(y;\theta)\,dy \tag{5.14}$$

Data Setup

For estimating the various inequality and poverty measures, we assume we have a sample $\boldsymbol{y}' = (y_1, y_2, \ldots., y_T)$ randomly drawn from $f(y;\theta)$, and grouped into N income classes $(x_0, x_1), (x_1, x_2), \ldots, (x_{N-1}, x_N)$ with $x_0 = 0$ and $x_N = \infty$. We denote the proportion of observations in the i-th group as c_i, mean income in the i-th group as $\overline{y}_i$, and mean income for the whole sample as $\overline{y}$. The income share for the i-th group is $s_i = c_i\overline{y}_i/\overline{y}$. Sometimes observations $\boldsymbol{c}' = (c_1, c_2, \ldots, c_N)$ and $\boldsymbol{s}' = (s_1, s_2, \ldots s_N)$ are available from one source and $\overline{y}$ is available from another source, in which case group mean incomes can be found from $\overline{y}_i = s_i\overline{y}/c_i$. In the next section, we describe various methods for estimating θ, given the observations $(c_i, s_i, \overline{y})$.

Estimation

The estimation methods that we review can be categorized according to the way in which the data are generated, and whether the group bounds $\boldsymbol{x}' = (x_0, x_1, \ldots, x_N)$ are known in addition to the observations on $(c_i, s_i, \overline{y})$. There are two ways in which the data can be generated. The group bounds $\boldsymbol{x}$ can be specified a priori, making the proportions of observations which fall into each group c_i, and the group means $\overline{y}_i$, the random variables. Alternatively, the c_i can be specified a priori, in which case the group bounds $\boldsymbol{x}$ are random variables, along with the group means $\overline{y}_i$. We consider estimation techniques for each of these cases in turn, noting the implications of known and unknown values for the group boundaries.

Estimation with Fixed x, Random c, Random $\overline{y}_i$

One approach for estimating θ when the group bounds $\boldsymbol{x}$ are known and the c_i are random is to maximize the likelihood function for the multinomial distribution. This approach uses information on $\boldsymbol{x}$ and $\boldsymbol{c}$, but does not utilize the information contained in $\boldsymbol{s}$ and $\overline{y}$. The log of the likelihood function is given by

$$L(\theta) \propto K + \sum_{i=1}^{N} c_i \ln\left[F(x_i; \theta) - F(x_{i-1}; \theta)\right] \tag{5.15}$$

where K is a constant.

In a series of papers (Griffiths & Hajargasht, 2015; Hajargasht & Griffiths, 2020; Hajargasht et al., 2012), three minimum distance (MD) estimators suitable for random c_i and $\overline{y}_i$ were introduced.[5] These estimators utilize information on $\boldsymbol{c}$, $\boldsymbol{s}$ and $\overline{y}$, and can be applied with or without knowledge of $\boldsymbol{x}$. When $\boldsymbol{x}$ is unknown it can be treated as a set of unknown parameters and estimated along with θ. The three estimators all have the same limiting distribution, but do not yield identical estimates. They are more efficient than the ML estimator from the multinomial likelihood

[5] The estimator in the first of these papers was described as a generalized method of moments estimator. Here, we use the term minimum distance estimator because it includes not only estimators that minimize the squared distance between sample and population moments, but also those that minimize the squared distance between sample quantities and their probability limits.

function where only information from c_i is utilized. To introduce the three estimators, we begin by noting the following:

$$\text{plim}\, c_i = F(x_i;\theta) - F(x_{i-1};\theta) = \lambda_i(\phi) - \lambda_{i-1}(\phi) \tag{5.16}$$

$$\text{plim}\, s_i = F^{(1)}(x_i;\theta) - F^{(1)}(x_{i-1};\theta) = \eta_i(\phi) - \eta_{i-1}(\phi) \tag{5.17}$$

where we write $\phi = (x, \theta)$ to accommodate the case where $\boldsymbol{x}$ is unobserved, making the unknown parameter vector equal to ϕ. If $\boldsymbol{x}$ is observed, we can proceed in the same way, utilizing the known $\boldsymbol{x}$ and treating θ as the unknown parameter vector.

MD Estimator 1

For the first MD estimator, we define

$$\tilde{y}_i = s_i\overline{y} = c_i\overline{y}_i \tag{5.18}$$

Since $\sum_{i=1}^{N} \tilde{y}_i = \sum_{i=1}^{N} c_i\overline{y}_i = \overline{y}$, we interpret $\tilde{y}_i$ as that part of mean income $\overline{y}$ that comes from the i-th group. Then, from (5.17) and (5.18),

$$\begin{aligned}\text{plim}\, \tilde{y}_i &= \text{plim}\, \bar{y}\, \text{plim}\, s_i \\ &= \mu\Big[F^{(1)}(x_i;\theta) - F^{(1)}(x_{i-1};\theta)\Big] \\ &= \mu\big[\eta_i(\phi) - \eta_{i-1}(\phi)\big] \end{aligned} \tag{5.19}$$

From (5.16) and (5.19), we can set up the MD estimator

$$\hat{\phi}_1 = \arg\min_{\phi} H_1(\phi)' W H_1(\phi) \tag{5.20}$$

where

$$H_1(\phi) = \begin{bmatrix} c_1 - [\lambda_1(\phi) - \lambda_0(\phi)] \\ \vdots \\ c_{N-1} - \big[\lambda_{N-1}(\phi) - \lambda_{N-2}(\phi)\big] \\ \tilde{y}_1 - \mu[\eta_1(\phi) - \eta_0(\phi)] \\ \vdots \\ \tilde{y}_N - \mu\big[\eta_N(\phi) - \eta_{N-1}(\phi)\big] \end{bmatrix} \tag{5.21}$$

and W is a weight matrix. Note that μ will also depend on ϕ, the exact function depending on the parametric pdf chosen for the income distribution. Also, $c_N - \left[\lambda_N(\phi) - \lambda_{N-1}(\phi)\right]$ has been omitted since having $\sum_{i=1}^{N} c_i = 1$ makes one of the c_i entries redundant.

A possible weight matrix, one suggested by Chotikapanich et al. (2007), is to set the diagonal elements of W as $w_i = 1/c_i^2$ for $i = 1, 2, \ldots, N-1$ and $w_{N-1+i} = 1/\tilde{y}_i^2$ for $i = 1, 2, \ldots, N$, and the off-diagonal elements to zero. With this setting $\hat{\phi}_1$ minimizes the sum of squares of percentage errors. This weight matrix, call it W_{CGR}, is a simple one, and it works well in practice, but it is not optimal; it does not lead to the most efficient estimator for ϕ. Hajargasht et al. (2012) show that the inverse of the optimal weight matrix is given by

$$W_1^{-1}(\phi) = \begin{bmatrix} D_1 & \begin{bmatrix} D_2 & 0_{N-1} \end{bmatrix} \\ \begin{bmatrix} D_2 \\ 0'_{N-1} \end{bmatrix} & D_3 \end{bmatrix} - \begin{bmatrix} A_1 & A_2 \\ A_2' & A_3 \end{bmatrix} \tag{5.22}$$

where 0_{N-1} is an $(N-1)-$ dimensional vector of zeros, and D_1, D_2 and D_3 are diagonal matrices. Their elements, and those of A_1, A_2 and A_3, are as follows.

$$[D_1]_{ii} = \lambda_i - \lambda_{i-1}, \qquad i = 1, 2, \ldots, N-1$$

$$[D_2]_{ii} = \mu(\eta_i - \eta_{i-1}), \qquad i = 1, 2, \ldots, N-1$$

$$[D_3]_{ii} = \mu^{(2)}(\psi_i - \psi_{i-1}), \qquad i = 1, 2, \ldots, N$$

$$[A_1]_{ij} = (\lambda_i - \lambda_{i-1})\left(\lambda_j - \lambda_{j-1}\right), \qquad i, j = 1, 2, \ldots, N-1$$

$$[A_2]_{ij} = (\lambda_i - \lambda_{i-1})\left(\eta_j - \eta_{j-1}\right) \qquad i = 1, 2, \ldots, N-1;\ j = 1, 2, \ldots, N$$

$$[A_3]_{ij} = (\eta_i - \eta_{i-1})\left(\eta_j - \eta_{j-1}\right), \qquad i, j = 1, 2, \ldots, N$$

All these quantities depend on the unknown parameter vector ϕ. To ease the notation, we have not made this dependence explicit. Note also that, through D_3, W will depend on the second moment $\mu^{(2)}$ and the second moment distribution function $\psi_i = F^{(2)}(x_i; \theta)$.

After inverting W_1^{-1} to find W_1, and simplifying, the objective function in (5.20) can be shown to be equal to

$$\begin{aligned} H_1(\phi)' W_1(\phi) H_1(\phi) &= \sum_{i=1}^{N} w_{1i}\left[c_i - (\lambda_i - \lambda_{i-1})\right]^2 \\ &+ \sum_{i=1}^{N} w_{2i}\left[\tilde{y}_i - \mu(\eta_i - \eta_{i-1})\right]^2 \\ &- 2\sum_{i=1}^{N} w_{3i}\left[c_i - (\lambda_i - \lambda_{i-1})\right]\left[\tilde{y}_i - \mu(\eta_i - \eta_{i-1})\right] \end{aligned} \tag{5.23}$$

where

$$w_{1i} = \frac{\mu^{(2)}(\psi_i - \psi_{i-1})}{v_i} \tag{5.24}$$

$$w_{2i} = \frac{(\lambda_i - \lambda_{i-1})}{v_i} \tag{5.25}$$

$$w_{3i} = \frac{\mu(\eta_i - \eta_{i-1})}{v_i} \tag{5.26}$$

and

$$v_i = \mu^{(2)}(\lambda_i - \lambda_{i-1})(\psi_i - \psi_{i-1}) - \mu^2(\eta_i - \eta_{i-1})^2$$

There are three possible ways to approach the problem of finding an estimate $\hat{\phi}$ that minimizes $H_1(\phi)' W_1(\phi) H_1(\phi)$:

1. A two-step estimator where first an estimate $\hat{\phi}_{CGR}$ is obtained using the weight matrix W_{CGR}, and then a second estimate $\hat{\phi}_{2-STEP}$ is obtained by minimizing $H_1(\phi)' W_1(\hat{\phi}_{CGR}) H_1(\phi)$.
2. An iterative estimator obtained by iterating the 2-step estimator until convergence is achieved.
3. A “continuous updating estimator” where the whole function in (5.23) is minimized with respect to ϕ.

These three estimators all have the same limiting distribution but can produce different estimates. Their asymptotic covariance matrix is

$$\operatorname{var}\left(\hat{\phi}_1\right) = \frac{1}{T}\left[\left(\frac{\partial H_1^{*\prime}}{\partial \phi}\right) W_1^* \left(\frac{\partial H_1^*}{\partial \phi^\prime}\right)\right]^{-1} \tag{5.27}$$

where H_1^* is a $(2N \times 1)$ vector obtained from H_1 by including $c_N - (\lambda_N - \lambda_{N-1})$ in the N-th position, and W_1^* is a $(2N \times 2N)$ matrix with 4 $(N \times N)$ diagonal blocks $D_{11},\ D_{12},\ D_{21} = D_{12}$ and D_{22}. The i-th diagonal elements of these matrices are w_{1i} for D_{11}, w_{2i} for D_{22} and $-w_{3i}$ for D_{12}. See Eqs. (5.24) to (5.26).

MD Estimator 2

The second MD estimator is that considered by Griffiths and Hajargasht (2015). It follows the same principles as the previous one, but it replaces $\tilde{y}_i$ by $\overline{y}_i$. To accommodate this replacement, we note that, from (5.16)–(5.18),

$$\begin{aligned} \operatorname{plim} \bar{y}_i &= \frac{\operatorname{plim} \bar{y} \operatorname{plim} s_i}{\operatorname{plim} c_i} \\ &= \frac{\mu(\eta_i - \eta_{i-1})}{\lambda_i - \lambda_{i-1}} \end{aligned}$$

In this case, the MD estimator can be written as

$$\hat{\phi}_2 = \arg\min_\phi H_2(\phi)^\prime W_2 H_2(\phi) \tag{5.28}$$

where

$$H_2(\phi) = \begin{bmatrix} c_1 - [\lambda_1 - \lambda_0] \\ \vdots \\ c_{N-1} - \left[\lambda_{N-1} - \lambda_{N-2}\right] \\ \overline{y}_1 - \dfrac{\mu[\eta_1 - \eta_0]}{\lambda_1 - \lambda_0} \\ \vdots \\ \overline{y}_N - \dfrac{\mu\left[\eta_N - \eta_{N-1}\right]}{\lambda_N - \lambda_{N-1}} \end{bmatrix} \tag{5.29}$$

and W_2 is a specified weight matrix. The weight matrix that is analogous to W_{CGR}, suggested for the previous estimator as a simple choice, or as a starting point for estimators that use an optimal weight matrix, is a diagonal matrix with elements $w_i = 1/c_i^2$ for $i = 1, 2, \ldots, N-1$, and $w_{i+N-1} = 1/\overline{y}_i^2$ for $i = 1, 2, \ldots, N$. Griffiths and Hajargasht (2015) show that the optimal weight matrix, for use with a 2-step, iterative or continuous updating estimator, is given by

$$W_2(\phi) = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \tag{5.30}$$

where

$$[E_1]_{ij} = \frac{\delta_{ij}}{\lambda_i - \lambda_{i-1}} + \frac{1}{\lambda_N - \lambda_{N-1}} \qquad i, j = 1,\ 2,\ \ldots,\ N-1 \tag{5.31}$$

$$[E_2]_{ij} = \frac{\delta_{ij}(\lambda_i - \lambda_{i-1})^3}{\mu^{(2)}(\lambda_i - \lambda_{i-1})(\psi_i - \psi_{i-1}) - \mu^2(\eta_i - \eta_{i-1})^2} \quad i, j = 1,\ 2,\ \ldots,\ N \tag{5.32}$$

and $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$. Using these results, the objective function can be simplified to

$$\begin{aligned} H_2(\phi)' W_2(\phi) H_2(\phi) &= \sum_{i=1}^{N} \frac{\left[c_i - (\lambda_i - \lambda_{i-1})\right]^2}{\lambda_i - \lambda_{i-1}} \\ &\quad + \sum_{i=1}^{N} [E_2]_{ii} \left(\bar{y}_i - \frac{\mu(\eta_i - \eta_{i-1})}{\lambda_i - \lambda_{i-1}} \right)^2 \end{aligned} \tag{5.33}$$

As before, $H_2(\phi)' W_2(\phi) H_2(\phi)$ can be minimized using a 2-step estimator, an iterative estimator or a continuous updating estimator. The weights are $1/(\lambda_i - \lambda_{i-1})$ for the first terms in (5.33) and $[E_2]_{ii}$ for the second. In contrast to the earlier formulation in (5.23), there are no cross product terms, making the minimization problem simpler and convergence easier to obtain. The large sample covariance matrix of an estimator $\hat{\phi}_2$ using an optimal weight matrix is

$$\operatorname{var}\left(\hat{\phi}_1\right) = \frac{1}{T} \left[\left(\frac{\partial H_1^{*'}}{\partial \phi} \right) W_1^* \left(\frac{\partial H_1^*}{\partial \phi'} \right) \right]^{-1} \tag{5.34}$$

where H_2^* is a $(2N \times 1)$ vector obtained from H_2 by including $c_N - (\lambda_N - \lambda_{N-1})$ in the N-th position, and W_2^* is a $(2N \times 2N)$ block-diagonal matrix with elements $1\big/(\lambda_i - \lambda_{i-1})$ in the first diagonal block and elements $[E_2]_{ii}$ in the second diagonal block.

MD Estimator 3

The third MD estimator that we describe is that considered by Hajargasht and Griffiths (2020). Its essential difference is that it considers cumulative population and income shares. To develop it, we begin by defining.

$$\hat{\lambda}_i = \sum_{j=1}^{i} c_j \text{ and } \hat{\eta}_i = \sum_{j=1}^{i} s_j \tag{5.35}$$

.

and recognizing that

$$\operatorname{plim} \hat{\lambda}_i = F(x_i;\theta) = \lambda_i(\phi) \tag{5.36}$$

$$\operatorname{plim} \bar{y}\,\hat{\eta}_i = \mu\, F^{(1)}(x_i;\theta) = \mu\,\eta_i(\phi) \tag{5.37}$$

Using (5.36) and (5.37), we can construct the MD estimator as

$$\hat{\phi}_3 = \arg\min_{\phi} H_3(\phi)' W_3 H_3(\phi) \tag{5.38}$$

where

$$H_3(\phi) = \begin{bmatrix} \hat{\lambda}_1 - \lambda_1 \\ \vdots \\ \hat{\lambda}_{N-1} - \lambda_{N-1} \\ \bar{y}\,\hat{\eta}_1 - \mu\,\eta_1 \\ \vdots \\ \bar{y}\,\hat{\eta}_{N-1} - \mu\,\eta_{N-1} \\ \bar{y} - \mu \end{bmatrix} \tag{5.39}$$

and W_3 is a pre-specified weight matrix. A simple weight matrix that can be used to simplify calculations or as a starting point for estimators that use an optimal weight matrix is a diagonal matrix with elements $w_i =$

$1\big/\hat{\lambda}_i^2$ for $i = 1, 2, \ldots, N-1$ and $w_{N-1+i} = 1\big/\left(\overline{y}\,\hat{\eta}_i\right)^2$ for $i = 1, 2, \ldots, N$. Hajargasht and Griffiths (2020) show that the optimal weight matrix is given by

$$W_3(\phi) = \begin{bmatrix} L_{11} & L_{12} \\ L_{12}' & L_{22} \end{bmatrix} \tag{5.40}$$

where

1. L_{11} is a $[(N-1) \times (N-1)]$ tri-diagonal matrix with the following nonzero elements:

$$[L_{11}]_{ii} = \frac{\mu^{(2)}(\psi_{i+1} - \psi_i)}{v_{i+1}} + \frac{\mu^{(2)}(\psi_i - \psi_{i-1})}{v_i} \quad i = 1, 2, \ldots, N-1$$

$$[L_{11}]_{ij} = \begin{cases} -\dfrac{\mu^{(2)}(\psi_i - \psi_{i-1})}{v_i} & i = 2,\ 3,\ \ldots,\ N-1;\ j = i-1 \\ -\dfrac{\mu^{(2)}\left(\psi_j - \psi_{j-1}\right)}{v_j} & j = 2,\ 3,\ \ldots,\ N-1;\ i = j-1 \end{cases} \tag{5.41}$$

2. L_{12} is a $[(N-1) \times N]$ matrix with the following nonzero elements:

$$[L_{12}]_{ii} = -\frac{\mu(\eta_{i+1} - \eta_i)}{v_{i+1}} - \frac{\mu(\eta_i - \eta_{i-1})}{v_i} \quad i = 1,\ 2,\ \ldots,\ N-1$$

$$[L_{12}]_{ij} = \begin{cases} \dfrac{\mu(\eta_i - \eta_{i-1})}{v_i} & i = 2,\ 3,\ \ldots,\ N-1;\ j = i-1 \\ \dfrac{\mu\left(\eta_j - \eta_{j-1}\right)}{v_j} & j = 2,\ 3,\ \ldots,\ N;\ i = j-1 \end{cases} \tag{5.42}$$

3. L_{22} is a $[N \times N]$ tri-diagonal matrix with the following nonzero elements

$$
\begin{aligned}
[L_{22}]_{ii} &= \frac{(\lambda_{i+1} - \lambda_i)}{v_{i+1}} + \frac{(\lambda_i - \lambda_{i-1})}{v_i} \quad i = 1, 2, \ldots, N-1 \\
[L_{22}]_{ij} &= \begin{cases} \dfrac{-(\lambda_i - \lambda_{i-1})}{v_i} & i = 2, 3, \ldots, N;\ j = i-1 \\ \dfrac{-\left(\lambda_j - \lambda_{j-1}\right)}{v_j} & j = 2, 3, \ldots, N;\ i = j-1 \end{cases} \\
[L_{22}]_{NN} &= \frac{-(\lambda_N - \lambda_{N-1})}{v_N}
\end{aligned}
\tag{5.43}
$$

As in the previous two cases, the objective function can be minimized using a two-step estimator, an iterative estimator or a continuous updating estimator. The asymptotic covariance matrix for $\hat{\phi}_3$, when using an optimal covariance matrix, is

$$
\operatorname{var}(\hat{\phi}_3) = \frac{1}{T}\left[\left(\frac{\partial H_3'}{\partial \phi}\right) W_3 \left(\frac{\partial H_3}{\partial \phi'}\right)\right]^{-1} \tag{5.44}
$$

A Quasi ML Estimator

Building on the work of Hilomi et al. (2008), Eckernkemper and Gribisch (2021), propose a quasi ML estimator. They combine the multinomial likelihood in Eq. (5.15) with a Gaussian approximation for the group means $\overline{y}_i$. Including the extra information means that estimation can proceed with or without knowledge of the group bounds, with these bounds being treated as parameters to be estimated when they are unknown. Let $T_i = c_i T$ be the number of observations in groups i. Each $\overline{y}_i$ is assumed to be $N\left(\tilde{\mu}_i, \tilde{\sigma}_i^2 / T_i\right)$ where the $\tilde{\mu}_i$ and the $\tilde{\sigma}_i^2$ are the means and variances of y from truncations $(x_{i-1} < y_i < x_i)$ of the originally specified distribution. That is,

$$
\tilde{\mu}_i = E(y | x_{i-1} < y < x_i) = \frac{\mu\left[\eta_i(\phi) - \eta_{i-1}(\phi)\right]}{\lambda_i(\phi) - \lambda_{i-1}(\phi)} \tag{5.45}
$$

and

$$\tilde{\sigma}_i^2 = \text{var}(y|x_{i-1} < y < x_i) = \frac{\mu^{(2)}\left[\psi_i(\phi) - \psi_{i-1}(\phi)\right]}{\lambda_i(\phi) - \lambda_{i-1}(\phi)} - \tilde{\mu}_i^2 \tag{5.46}$$

Using these results, the log of the likelihood function can be written as

$$L(\phi) \propto K_1 + \sum_{i=1}^{N} \left\{ c_i \ln\left[\lambda_i(\phi) - \lambda_{i-1}(\phi)\right] - \ln \tilde{\sigma}_i - \frac{c_i}{2\tilde{\sigma}_i^2}(\overline{y} - \tilde{\mu}_i)^2 \right\} \tag{5.47}$$

Eckernkemper and Gribisch (2021) show that the estimator for ϕ that maximizes $L(\phi)$ is consistent and that the covariance matrix of its limiting distribution is the same as that for MD estimators 1 and 2.

Estimation with Fixed c, Random x, Random $\overline{y}_i$

In this case, the observations are grouped such that the proportion of observations in each group is pre-specified. Examples are 10 groups with 10% of the observations in each group or 20 groups with 5% of the observations in each group. This setup implies the proportion c_i are fixed (non-random) and the sample group boundaries $\boldsymbol{x}$ as well as the average cumulative incomes $\overline{y}_i$ are random variables. Let $y_{[1]}, y_{[2]}, \ldots, y_{[T]}$ be the order statistics obtained by arranging the original observations $\boldsymbol{y}$ in ascending order. An estimate for a group bound x_i is the largest order statistic in the i-th group, $\hat{x}_i = y_{[\hat{\lambda}_i T]}$. If the $\hat{x}_i$ are observed, estimation can use both the $\hat{x}_i$ and the $\overline{y}_i$; if the $\hat{x}_i$ are unobserved, then only the information in $\overline{y}_i$ can be utilized. We consider MD and ML estimation for both these cases. MD estimation with unobserved $\hat{x}_i$ corresponds to Lorenz curve estimation which has attracted a great deal of attention in the literature. See, for example, Chotikapanich (2008). A Lorenz curve implied by a specific income distribution is defined by Eq. (5.4). An alternative is to start with a specific parametric Lorenz curve in which case the corresponding income distribution is defined via the quantile function in (5.5). A problem with the latter approach is that the income distributions corresponding to some Lorenz curves are not defined for all values of y.

MD Estimation

The MD estimators that we consider are those proposed by Hajargasht and Griffiths (2020). Suppose, in the first instance, that the $\hat{x}_i$ are observed. To use this information in an MD estimator, we recognize that[6]

$$\text{plim}\,\hat{x}_i = F^{-1}\left(\hat{\lambda}_i;\theta\right) \tag{5.48}$$

To use information on the income shares, we use the cumulative shares $\hat{\eta}_i$ multiplied by mean income $\overline{y}$, in line with MD estimator 3 for the random $\boldsymbol{c}$ case. One difference, however, is that we express its probability limit in terms of the non-random $\boldsymbol{c}$, instead of $\boldsymbol{x}$, which is now a random variable. That is,

$$\text{plim}\,\bar{y}\,\hat{\eta}_i = \mu\, F^{(1)}\left(F^{-1}(\hat{\lambda};\theta);\theta\right) \tag{5.49}$$

To set up the MD estimator, it is convenient to define notation for a generalized Lorenz curve which can be written as

$$\mu\,\eta = G(\lambda;\theta) = \mu\,L(\lambda;\theta) = \mu\,F^{(1)}\left(F^{-1}(\lambda;\theta);\theta\right) \tag{5.50}$$

Then, from (5.48)–(5.50), we can set up the following MD estimator,

$$\hat{\theta}_4 = \arg\min_\theta H_4'(\theta)\,W_4\,H_4(\theta) \tag{5.51}$$

[6] To avoid introducing more notation to what is already a very substantial amount, we will continue to use $\hat{\lambda}_i$ to denote the observed cumulative proportion of population, despite the fact that, in the current context, it is a non-random fixed quantity.

where

$$H_4(\theta) = \begin{bmatrix} \hat{x}_1 - F^{-1}(\hat{\lambda}_1; \theta) \\ \hat{x}_2 - F^{-1}(\hat{\lambda}_2; \theta) \\ \vdots \\ \hat{x}_{N-1} - F^{-1}(\hat{\lambda}_{N-1}; \theta) \\ \overline{y}\,\hat{\eta}_1 - G(\hat{\lambda}_1; \theta) \\ \vdots \\ \overline{y}\,\hat{\eta}_{N-1} - G(\hat{\lambda}_{N-1}; \theta) \\ \overline{y} - \mu \end{bmatrix} \tag{5.52}$$

and W_4 is a suitable chosen weight matrix. It can be shown that the optimal weight matrix is given by

$$W_4(\theta) = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{bmatrix}^{-1} \tag{5.53}$$

where

$$\underset{[(N-1)\times(N-1)]}{[\Omega_{11}]_{ij}} = \begin{cases} \dfrac{\hat{\lambda}_i(1-\hat{\lambda}_j)}{f(\hat{x}_i)f(\hat{x}_j)} & i \le j \\ \dfrac{\hat{\lambda}_j(1-\hat{\lambda}_i)}{f(\hat{x}_i)f(\hat{x}_j)} & j \le i \end{cases} \tag{5.54}$$

$$\underset{[N\times N]}{[\Omega_{22}]_{ij}} = \begin{cases} \mu^{(2)}\psi_i + \left[\hat{\lambda}_i\hat{x}_i - G(\hat{\lambda}_i)\right]\left[\hat{x}_j - \hat{\lambda}_j\hat{x}_j + G(\hat{\lambda}_j)\right] \\ \qquad - \hat{x}_i G(\hat{\lambda}_i) \quad i \le j \\ \mu^{(2)}\psi_j + \left[\hat{\lambda}_j\hat{x}_j - G(\hat{\lambda}_j)\right]\left[\hat{x}_i - \hat{\lambda}_i\hat{x}_i + G(\hat{\lambda}_i)\right] \\ \qquad - \hat{x}_j G(\hat{\lambda}_j) \quad j \le i \end{cases} \tag{5.55}$$

$$\underset{[(N-1)\times N]}{[\Omega_{12}]_{ij}} = \begin{cases} \dfrac{\hat{\lambda}_i[G(\hat{\lambda}_j) - \hat{x}_j\hat{\lambda}_j + \hat{x}_j] - G(\hat{\lambda}_i)}{f(\hat{x}_i)} & i \le j \\ \dfrac{[\hat{\lambda}_i - 1][G(\hat{\lambda}_j) - \hat{x}_j\hat{\lambda}_j]}{f(\hat{x}_i)} & j \le i \end{cases} \tag{5.56}$$

The covariance matrix for the limiting distribution of $\hat{\theta}_4$ is

$$\mathrm{var}(\hat{\theta}_4) = \frac{1}{T}\left(\frac{\partial H_4'}{\partial \theta} W_4 \frac{\partial H_4}{\partial \theta'}\right)^{-1} \quad (5.57)$$

When there are a large number of groups, the matrix inversion in (5.53) can be computationally demanding. Hajargasht and Griffiths (2020) show how W_4^{-1} can be derived from W_3^{-1} which has computationally convenient tri-diagonal blocks. They also demonstrate that, if the groupings for this set up are equivalent to those for the MD3 setup in the sense that, a priori, $x_i = F^{-1}(\lambda_i; \theta)$, then the asymptotic covariance matrices for $\hat{\theta}_3$ and $\hat{\theta}_4$ are identical.

Minimizing (5.51) to find an estimate $\hat{\theta}_4$ can proceed using one of the three algorithms described in section Estimation with Fixed $\boldsymbol{x}$, Random $\boldsymbol{c}$, Random $\overline{\boldsymbol{y}}_{\boldsymbol{i}}$. However, there are two requirements which will not always be met: estimates of the bounds $\hat{x}_i = y_{[\hat{\lambda}_i T]}$ must be observed and the cdf must be invertible, either algebraically or computationally, so that quantiles $F^{-1}(\hat{\lambda}_i; \theta)$ can be found. Note that $F^{-1}(\hat{\lambda}_i; \theta)$ appears not only in the first $(N-1)$ elements of H_4 but also in the next $(N-1)$ elements that involve the generalized Lorenz curve $G(\lambda; \theta) = \mu\, F^{(1)}\big(F^{-1}(\lambda; \theta); \theta\big)$. One way to overcome non-invertibility of the cdf is to replace the assumption of a parametric income distribution with an assumption of a parametric Lorenz curve. Doing so overcomes the problem for the second set of elements in H_4, and relationships between the generalized Lorenz curve and the quantile function—see Hajargasht and Griffiths (2020)—can be exploited to obtain the first set of elements in H_4.

When the $\hat{x}_i$ are unobserved, estimation can proceed using the last N elements in H_4, with calculations made from an assumed income distribution if the cdf is invertible, or from an assumed Lorenz curve if the cdf is not invertible. This last approach is that most closely aligned with suggestions for Lorenz curve estimation which have appeared in the literature.[7] Earlier suggestions are sub-optimal in the sense that they do not use the best weighting matrix. Details can be found in Hajargasht and Griffiths (2020).

[7] See Chotikapanich (2008) for access to this literature.

ML Estimation

ML estimation of θ for fixed c_i, and random x_i and $\overline{y}_i$ was considered by Eckernkemper and Gribisch (2021). Recognizing that the joint density for the group bounds and group means can be written as

$$f(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N, \hat{\boldsymbol{x}}) = f(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N | \hat{\boldsymbol{x}}) f(\hat{x}_1) \\ f(\hat{x}_2 | \hat{x}_1) \ldots f(\hat{x}_{N-1} | \hat{x}_{N-2}) \tag{5.58}$$

they set up a likelihood function that uses distribution theory for order statistics for $f(\hat{\boldsymbol{x}})$ and a Gaussian approximation for $f(\overline{y}_i | \hat{x}_i, \hat{x}_{i-1})$. Using results in David and Nagaraji (2003), the conditional means and variances for the $\overline{y}_i$ can be written as

$$\mu_i = E(\overline{y}_i | \hat{x}_i, \hat{x}_{i-1}) = \frac{T_i - 1}{T_i} \tilde{\mu}_i + \frac{\hat{x}_i}{T_i} \tag{5.59}$$

and

$$\sigma_i^2 = \text{var}(\overline{y}_i | \hat{x}_i, \hat{x}_{i-1}) = \frac{T_i - 1}{T_i^2} \tilde{\sigma}_i^2 \tag{5.60}$$

The log-likelihood is

$$\begin{aligned} L(\theta) = K_2 - \frac{1}{2}\left[\ln \tilde{\sigma}_N^2 + \frac{T_N}{\tilde{\sigma}_N^2}(\bar{y}_N - \tilde{\mu}_N)^2\right] + T_N \ln\left[1 - F(\hat{x}_{N-1}; \theta)\right] \\ + \sum_{i=1}^{N-1} \left\{ \begin{array}{l} -\frac{1}{2}\left[\ln \sigma_i^2 + \left(\frac{\bar{y}_i - \mu_i}{\sigma_i}\right)^2\right] \\ + (T_i - 1)\ln\left[F(\hat{x}_i; \theta) - F(\hat{x}_{i-1}; \theta)\right] + \ln f(\hat{x}_i; \theta) \end{array} \right\} \end{aligned} \tag{5.61}$$

where K_2 is a constant. The estimator $\hat{\theta}_5$ that minimizes (5.61) can be interpreted as a quasi ML estimator. Eckernkemper and Gribisch (2021) establish its asymptotic covariance matrix as

$$\text{var}(\hat{\theta}_5) = \frac{1}{T}\left[\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \theta} \frac{\partial \mu_i}{\partial \theta'}\right)\left(\frac{c_i}{\tilde{\sigma}_i^2}\right)\right]^{-1} \tag{5.62}$$

One difference between the estimator $\hat{\theta}_5$ and those estimators considered in the earlier sections is that it requires knowledge of the sample size T, from which the number of observations in each group can be found from $T_i = c_i T$. All estimators require knowledge of T to compute standard errors, but knowledge of the proportions c_i, without knowledge of T, is sufficient for the earlier estimators for θ and ϕ to be employed.

For ML estimation of θ when the $\hat{x}_i$ are not observed, Eckernkemper and Gribisch (2021) integrate out the $\hat{x}_i$ from the likelihood in (5.61) to obtain the following log-likelihood

$$L(\theta) = K_3 - \frac{1}{2}\left[\log|\Xi| + T\left(\overline{\mathbf{y}} - \boldsymbol{\mu}^*\right)'\Xi^{-1}\left(\overline{\mathbf{y}} - \boldsymbol{\mu}^*\right)\right] \tag{5.63}$$

where K_3 is a constant, $\overline{\mathbf{y}}' = (\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_N)$, $\boldsymbol{\mu}^*$ is an $(N \times 1)$ vector with i-th element equal to

$$\mu_i^* = \frac{1}{c_i}\left[G\left(\hat{\lambda}; \theta\right) - G\left(\hat{\lambda}_{i-1}; \theta\right)\right] \tag{5.64}$$

and $\Xi = DB\,\Omega_{22}^* B'D$ where $D = \operatorname{diag}\left(c_1^{-1}, c_2^{-1}, \ldots, c_N^{-1}\right)$, $[B]_{ii} = 1$, $[B]_{ij} = -1$ for $i = j + 1$, $j = 1, 2, \ldots, N - 1$, and zero elsewhere, and Ω_{22}^* is equal to Ω_{22} defined in (5.55), but with $\hat{x}_i$ and $\hat{x}_j$ replaced by $F^{-1}\left(\hat{\lambda}_i; \theta\right)$ and $F^{-1}\left(\hat{\lambda}_j; \theta\right)$, respectively. The asymptotic covariance matrix for the estimator $\hat{\theta}_6$ obtained by maximizing (5.61) is

$$\operatorname{var}(\hat{\theta}_6) = \frac{1}{T}\left[\partial\mu^{*'}\,\Xi^{-1}\frac{\partial\mu^*}{\partial\theta'}\right] \tag{5.65}$$

Specification of Distributions, Inequality and Poverty Measures

To implement the estimation methods described in section Estimation, a specific parametric distribution has to be specified and we need its moments, its pdf, cdf, fmdf and smdf. This information is provided in Table 5.1 for several popular income distributions. Once the parameters of a chosen distribution have been estimated, estimates for inequality and poverty incidence are frequently of interest. In Table 5.2, we provide expressions that can be used to compute inequality estimates from the estimates of the parameters. Expressions for the poverty estimates were

Table 5.1 Probability distributions, distribution functions and moments

Distribution	*pdf*	*cdf/fmdf/smdf* [a]	*Moments*
Pareto	$f(y;\alpha,y_0)=\frac{\alpha y_0^{\alpha}}{y^{\alpha+1}}$ $y\geq y_0>0$	$F^{(k)}(y;\alpha,y_0)=1-\left(\frac{y_0}{y}\right)^{\alpha-k}$ $\alpha>k$	$\mu^{(k)}=\frac{\alpha y_0^k}{\alpha-k}$ $\alpha>k$
Lognormal[b]	$f(y;\beta,\sigma)=$ $\frac{1}{\sigma y\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\ln y-\mu)^2\right\}$	$F^{(k)}(y;\beta,\sigma)=\Phi\left(\frac{\ln(y)-\beta-k\sigma^2}{\sigma}\right)$	$\mu^{(k)}=$ $\exp\left\{k\beta+\frac{k^2\sigma^2}{2}\right\}$
Pareto-lognormal	$f(y;\alpha,\beta,\sigma)=$ $\frac{\alpha}{y^{\alpha+1}}\exp\left\{\frac{\alpha^2\sigma^2}{2}+\alpha\beta\right\}\Phi\left(\frac{\ln y-\beta-\alpha\sigma^2}{\sigma}\right)$	$F^{(k)}(y;\alpha,\beta,\sigma)=\Phi\left(\frac{\ln y-\beta-k\sigma^2}{\sigma}\right)-$ $y^{k-\alpha}\exp\left\{\frac{\sigma^2}{2}(\alpha^2-k^2)+\beta(\alpha-k)\right\}\Phi\left(\frac{\ln y-\beta-\alpha\sigma^2}{\sigma}\right)$ $\alpha>k$	$\mu^{(k)}=$ $\frac{\alpha}{\alpha-k}\exp\left\{k\beta+\frac{k^2\sigma^2}{2}\right\}$ $\alpha>k$
GB2[c]	$f(y;a,b,p,q)=$ $\frac{ay^{ap-1}}{b^{ap}B(p,q)\left[1+\left(\frac{y}{b}\right)^a\right]^{p+q}}$	$F^{(k)}(y;a,b,p,q)=B_u\left(p+\frac{k}{a},q-\frac{k}{a}\right)$ $u=\frac{(y/b)^a}{1+(y/b)^a}\quad k<qa$	$\mu^{(k)}=$ $\frac{b^k B\left(p+\frac{k}{a},q-\frac{k}{a}\right)}{B(p,q)}$ $k<qa$
Singh-Maddala[d] ($p=1$)	$f(y;a,b,q)=\frac{aqy^{a-1}}{b^a\left[1+\left(\frac{y}{b}\right)^a\right]^{1+q}}$	$F(y;a,b,q)=1-\left[1+\left(\frac{y}{b}\right)^a\right]^{-q}$ $F^{(k)}(y;a,b,q)=B_u\left(1+\frac{k}{a},q-\frac{k}{a}\right)$ $u=\frac{(y/b)^a}{1+(y/b)^a}\quad k<aq$	$\mu^{(k)}=$ $\frac{b^k\Gamma\left(1+\frac{k}{a}\right)\Gamma\left(q-\frac{k}{a}\right)}{\Gamma(q)}$ $k<aq$
Beta-2 ($a=1$)	$f(y;b,p,q)=\frac{y^{p-1}}{b^pB(p,q)\left[1+\frac{y}{b}\right]^{p+q}}$	$F^{(k)}(y;b,p,q)=B_u(p+k,q-k)$ $u=\frac{y/b}{1+y/b}\quad k<q$	$\mu=\frac{bp}{q-1}$ $\mu^{(2)}=\frac{bp(p+1)}{(q-1)(q-2)}$

Distribution	*pdf*	*cdf/fmdf/smdf* [a]	*Moments*
Dagum $(q = 1)$	$f(y;a,b,p) = \frac{apy^{ap-1}}{b^{ap}\left[1+\left(\frac{y}{b}\right)^a\right]^{p+1}}$	$F(y;a,b,p) = \left[1+\left(\frac{y}{b}\right)^{-a}\right]^{-p}$ $F^{(k)}(y;a,b,p) = B_u\left(p+\frac{k}{a}, 1-\frac{k}{a}\right)$ $u = \frac{(y/b)^a}{1+(y/b)^a}$ $k < a$	$\mu^{(k)} = \frac{b^k\Gamma\left(p+\frac{k}{a}\right)\Gamma\left(1-\frac{k}{a}\right)}{\Gamma(p)}$ $k < a$
Generalized gamma$^{(5)}$	$f(y;a,b,p) = \frac{ay^{ap-1}\exp\left\{-\left(\frac{y}{b}\right)^a\right\}}{b^{ap}\Gamma(p)}$	$F^{(k)}(y;a,b,p) = \Gamma_u\left(p+\frac{k}{a}\right)$ $u = \left(\frac{y}{b}\right)^a$	$\mu^{(k)} = \frac{b^k\Gamma\left(p+\frac{k}{a}\right)}{\Gamma(p)}$
Gamma $(a = 1)$	$f(y;b,p) = \frac{y^{p-1}\exp\left\{-\frac{y}{b}\right\}}{b^p\Gamma(p)}$	$F^{(k)}(y;b,p) = \Gamma_u(p+k)$ $u = \frac{y}{b}$	$\mu = bp$ $\mu^{(2)} = b^2p(p+1)$

Notes Kleiber and Kotz (2003) is an excellent source for many of the results in this table

[a] Of interest are $k = 0$ (cdf = cumulative distribution function), $k = 1$ (fmdf = first moment distribution function), and $k = 2$ (smdf = second moment distribution function

[b] $\Phi(x)$ is the standard normal cdf evaluated at x

[c] GB2 is the generalized beta distribution of the 2nd kind. $B(p,q)$ is the beta function evaluated at (p,q). $B_u(p,q)$ is the standard beta cdf, with parameters (p,q), evaluated at u

[d] $\Gamma(p)$ is the gamma function evaluation at p

[e] $\Gamma_u(p)$ is the standard gamma cdf, with parameter p, evaluated at u

given in section Poverty Measures, with the exception of the Watts Index whose expressions we have tabulated in Table 5.3.

Simple Recipes for Two Distributions

In some instances, where large scale projects involving many countries and many time periods are being undertaken, it may be prudent to use estimation techniques which are relatively simple. In this section, we consider two estimation techniques that fall into this category—one for the lognormal distribution and one for the Pareto-lognormal distribution.

Lognormal Distribution

In the previous section, we indicated that the Gini coefficient for the lognormal distribution is $g = 2\Phi\left(\sigma \big/ \sqrt{2}\right) - 1$ and its mean is $\mu = \exp\left\{\beta + \sigma^2 \big/ 2\right\}$. Using grouped data the Gini coefficient can be estimated from

$$\hat{g} = \sum_{i=1}^{N-1} \hat{\eta}_{i+1}\hat{\lambda}_i - \sum_{i=1}^{N-1} \hat{\eta}_i\hat{\lambda}_{i+1} \tag{5.66}$$

and the mean can be estimated using $\overline{y}$,

$$\hat{\mu} = \overline{y} = \exp\left\{\hat{\beta} + \frac{\hat{\sigma}^2}{2}\right\} \tag{5.67}$$

Utilizing these two equations and the expression for the Gini coefficient yields the parameter estimates.

$$\hat{\sigma} = \sqrt{2}\Phi^{-1}\left(\frac{g+1}{2}\right) \tag{5.68}$$

$$\hat{\beta} = \ln(\overline{y}) - \frac{\hat{\sigma}^2}{2} \tag{5.69}$$

This approach was adopted by Chotikapanich et al. (1997).

Pareto-Lognormal Distribution

For the Pareto-lognormal distribution, we can estimate the Theil inequality measures from the grouped data, and then use these estimates,

Table 5.2 Inequality measures[a]

Distribution	*Gini coefficient*	*Theil index (0)*
Pareto	$g = \frac{1}{2\alpha-1}$	$T_0 = -\frac{1}{\alpha} + \ln\left(\frac{\alpha}{\alpha-1}\right)$
Lognormal	$g = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$	$T_0 = \frac{\sigma^2}{2}$
Pareto-lognormal	$g = \frac{2\exp\left\{\alpha(\alpha-1)\sigma^2\right\}}{2\alpha-1}\Phi\left(\frac{(1-2\alpha)\sigma}{\sqrt{2}}\right) + 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$	$T_0 = \ln\left(\frac{\alpha}{\alpha-1}\right) - \frac{1}{\alpha} + \frac{\sigma^2}{2}$
GB2[b]	See Kleiber and Kotz (2003, p. 193)[c]	$T_0 = \ln\left(\frac{\mu}{b}\right) - \frac{\Psi(p)-\Psi(q)}{a}$
Beta-2 ($a = 1$)	$g = \frac{2B(2p,2q-1)}{pB^2(p,q)}$	$T_0 = \ln\left(\frac{p}{q-1}\right) - [\Psi(p) - \Psi(q)]$
Singh-Maddala ($p = 1$)	$g = 1 - \frac{\Gamma(q)\Gamma\left(2q-\frac{1}{a}\right)}{\Gamma\left(q-\frac{1}{a}\right)\Gamma(2q)}$	$T_0 = \ln\left(\frac{\mu}{b}\right) - \frac{\Psi(1)-\Psi(q)}{a}$
Dagum ($q = 1$)	$g = \frac{\Gamma(p)\Gamma\left(2p+\frac{1}{a}\right)}{\Gamma\left(p+\frac{1}{a}\right)\Gamma(2p)} - 1$	$T_0 = \ln\left(\frac{\mu}{b}\right) - \frac{\Psi(p)-\Psi(1)}{a}$
Generalized Gamma	See Kleiber and Kotz (2003, p.155)[d]	$T_0 = \ln\left(\frac{\Gamma\left(p+\frac{1}{a}\right)}{\Gamma(p)}\right) - \frac{\Psi(p)}{p}$

(continued)

Table 5.2 (continued)

Distribution	*Gini coefficient*	*Theil index (0)*
Gamma ($a = 1$)	$g = \frac{\Gamma\left(p+\frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma(p+1)}$	$T_0 = \ln p - \frac{\Psi(p)}{p}$

Distribution	*Theil index (1)*	*Pietra index*
Pareto	$T_1 = \frac{1}{\alpha-1} - \ln\left(\frac{\alpha}{\alpha-1}\right)$	$P = \frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha}}$
Lognormal	$T_1 = \frac{\sigma^2}{2}$	$P = 2\Phi\left(\frac{\sigma}{2}\right) - 1$
Pareto-lognormal	$T_1 = \ln\left(\frac{\alpha-1}{\alpha}\right) + \frac{1}{\alpha-1} + \frac{\sigma^2}{2}$	$P = \Phi\left(u + \frac{\sigma}{2}\right) - \Phi\left(u - \frac{\sigma}{2}\right) + \frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha}} \exp\left\{\frac{\alpha\sigma^2}{2}(\alpha-1)\right\}\Phi\left(u + \sigma\left(\frac{1}{2} - \alpha\right)\right)$ $u = \frac{1}{\sigma}\ln\left(\frac{\alpha}{\alpha-1}\right)$
GB2[b]	$T_1 = \frac{\Psi\left(p+\frac{1}{a}\right) - \Psi\left(q-\frac{1}{a}\right)}{a} + \ln\left(\frac{b}{\mu}\right)$	$P = B_u(p,q) - B_u\left(p + \frac{1}{a}, q - \frac{1}{a}\right)$ $u = \frac{(\mu/b)^a}{1+(\mu/b)^a}$
Beta-2 ($a = 1$)	$T_1 = \Psi(p+1) - \Psi(q-1) + \ln\left(\frac{q-1}{p}\right)$	$P = \frac{u^p(1-u)^{q-1}}{pB(p,q)}$ $u = \frac{\mu/b}{1+\mu/b}$

Distribution	*Theil index (1)*	*Pietra index*
Singh-Maddala ($p = 1$)	$T_1 = \frac{\Psi\left(1+\frac{1}{a}\right)-\Psi\left(q-\frac{1}{a}\right)}{a} + \ln\left(\frac{b}{\mu}\right)$	$P = 1 - \left[1 + \left(\frac{\mu}{b}\right)^a\right]^{-q} - B_u\left(1 + \frac{1}{a}, q - \frac{1}{a}\right)$ $u = \frac{(\mu/b)^a}{1+(\mu/b)^a}$
Dagum ($q = 1$)	$T_1 = \frac{\Psi\left(p+\frac{1}{a}\right)-\Psi\left(1-\frac{1}{a}\right)}{a} + \ln\left(\frac{b}{\mu}\right)$	$P = \left[1 + \left(\frac{\mu}{b}\right)^{-a}\right]^{-p} - B_u\left(p + \frac{1}{a}, 1 - \frac{1}{a}\right)$ $u = \frac{(\mu/b)^a}{1+(\mu/b)^a}$
Generalized Gamma	$T_1 = \frac{\Psi\left(p+\frac{1}{a}\right)}{a} - \ln\left(\frac{\Gamma\left(p+\frac{1}{a}\right)}{\Gamma(p)}\right)$	$P = \Gamma_u(p) - \Gamma_u\left(p + \frac{1}{a}\right)$ $u = \left(\frac{y}{b}\right)^a$
Gamma ($a = 1$)	$T_1 = \frac{1}{p} + \Psi(p) - \ln p$	$P = \frac{p^p \exp\{-p\}}{\Gamma(p+1)}$

Notes [a]Many of these measures can be found in Kleiber and Kotz (2003). Sarabia et al (2017) has a comprehensive list of the Theil measures. Sarabia and Jordá (2014) consider the Pietra coefficient in detail

[b]$\Psi(p)$ is the digamma function $\Psi(p) = d \ln \Gamma(p) / dp$.

[c]The expression in Kleiber and Kotz involves lengthly hypergeometric functions. It also contains a typographical error; the first p in the denominator should be deleted. Computationally, we have found that evaluating the Gini coefficient via numerical integration is easier than computing values for the hypergeometric functions

[d]Like the GB2, this Gini expression involves hypergeometric functions. It may be preferable to evaluate the Gini integral numerically

Table 5.3 Watts poverty indices for selected distributions

Distribution	*Index*
Pareto	$WI = \ln\left(\frac{z}{y_0}\right) + \left(\frac{\alpha}{\alpha-1}\right)\left(\frac{1}{y_0}\right)\left[1 - \left(\frac{y_0}{z}\right)^{\alpha+1}\right]$
Lognormal	$WI = \sigma[u\Phi(u) + \phi(u)] \quad u = \frac{\ln z - \beta}{\sigma}$
Pareto-lognormal	$WI = -\frac{1}{\alpha}F(z;\alpha,\beta,\sigma) + \sigma[u\Phi(u) + \phi(u)] \quad u = \frac{\ln z - \beta}{\sigma}$
GB2[a]	$WI = \ln\left(\frac{z}{b}\right)B_u(p,q) -$ $\frac{1}{a}\{D_p B_u(p,q) - D_q B_u(p,q) + B_u(p,q)[\Psi(p) - \Psi(q)]\}$ $u = \frac{(z/b)^a}{1+(z/b)^a}$
Beta-2 ($a = 1$)	$WI = \ln\left(\frac{z}{b}\right)B_u(p,q) + D_q B_u(p,q) - D_p B_u(p,q) -$ $B_u(p,q)[\Psi(p) - \Psi(q)]$ $u = \frac{z/b}{1+z/b}$
Singh-Maddala ($p = 1$)	$WI = \ln\left(\frac{z}{b}\right)F(z;a,b,q) +$ $\frac{1}{a}\left\{\left[1 + \left(\frac{z}{b}\right)^a\right]^{-q}\ln\left[1 + \left(\frac{z}{b}\right)^a\right] - F(z;a,b,q)[\Psi(1) - \Psi(q)]\right\}$
Dagum ($q = 1$)	$WI = \ln\left(\frac{z}{b}\right)F(z;a,b,p) +$ $\frac{1}{a}\left\{\left[1 + \left(\frac{z}{b}\right)^{-a}\right]^{-p}\ln\left[1 + \left(\frac{z}{b}\right)^{-a}\right] - F(z;a,b,p)[\Psi(p) - \Psi(1)]\right\}$
Generalized Gamma[b]	$WI = \ln\left(\frac{z}{b}\right)\Gamma_u(p) - \frac{1}{a}\left[D_p\Gamma_u(p) + \Gamma_u(p)\Psi(p)\right]$ $u = \left(\frac{z}{b}\right)^a$
Gamma ($a = 1$)	$WI = \ln(u)\Gamma_u(p) - D_p\Gamma_u(p) + \Gamma_u(p)\Psi(p)$ $u = \frac{z}{b}$

Notes [a] $D_p B_u(p,q)$ and $D_q B_u(p,q)$ are the derivatives of the beta cdf $B_u(p,q)$ with respect to p and q respectively. These derivatives are available in some software such as EViews. A derivation of the expression can be found in Chotikapanich et al. (2013)
[b] $D_p\Gamma_u(p)$ is the derivative of the gamma cdf $\Gamma_u(p)$ with respect to p. It too is available in software such as EViews. Derivation of the result uses a similar approach to that for the GB2

along with sample mean income to estimate the parameters. Working in this direction, the grouped-data sample estimates are

$$\hat{T}_1 = \sum_{i=1}^{N} c_i \left(\frac{\overline{y}_i}{\overline{y}}\right) \ln\left(\frac{\overline{y}_i}{\overline{y}}\right) \tag{5.70}$$

$$\hat{T}_0 = \sum_{i=1}^{N} c_i \ln\left(\frac{\overline{y}}{\overline{y}_i}\right) \tag{5.71}$$

$$\hat{\mu} = \overline{y} \tag{5.72}$$

The corresponding quantities in terms of the parameters of the Pareto-lognormal distribution are

$$T_1 = \frac{1}{\alpha - 1} + \frac{\sigma^2}{2} + \ln\left(\frac{\alpha - 1}{\alpha}\right) \tag{5.73}$$

$$T_0 = -\frac{1}{\alpha} + \frac{\sigma^2}{2} - \ln\left(\frac{\alpha - 1}{\alpha}\right) \tag{5.74}$$

$$\mu = \frac{\alpha}{\alpha - 1} \exp\left\{\beta + \frac{\sigma^2}{2}\right\} \tag{5.75}$$

Assuming the mean exists ($\alpha > 1$), from (5.70)–(5.75) we can retrieve parameter estimates using the following three steps:

1. Find $\hat{\alpha}$ as the Solution to the Equation

$$\frac{2\hat{\alpha} - 1}{\hat{\alpha}(\hat{\alpha} - 1)} + 2\ln\left(\frac{\hat{\alpha} - 1}{\hat{\alpha}}\right) = \hat{T}_1 - \hat{T}_0 \tag{5.76}$$

2. Find $\hat{\sigma}^2$ from

$$\hat{\sigma}^2 = \hat{T}_1 + \hat{T}_0 - \frac{1}{\hat{\alpha}(\hat{\alpha} - 1)} \tag{5.77}$$

3. Find $\hat{\beta}$ from

$$\hat{\beta} = \ln(\overline{y}) + \ln\left(\frac{\hat{\alpha} - 1}{\hat{\alpha}}\right) - \frac{\hat{\sigma}^2}{2} \tag{5.78}$$

Concluding Remarks

Inequality and poverty, both nationally and globally, continue to be two of the most pressing issues facing today's society. Accurate measurement of inequality and poverty involves a multitude of non-trivial considerations including reliable data collection, specification of purchasing power parities and definition of a suitable poverty line. We have focused on a further consideration, how to model and estimate income distributions, and how to estimate inequality and poverty from the parameters of those income distributions, when using grouped data. Single observations are becoming increasingly available and their use is preferred to the use of grouped data if resources are adequate for doing so. However, countries and time periods for which only grouped data are available are still prevalent, and it can be advantageous to use grouped data for large scale regional and global projects. Our objective has been to summarize available techniques in a convenient form for researchers working along these lines.

References

Chotikapanich, D., Valenzuela, M. R., & Rao, D. S. P. (1997). Global and regional inequality in the distribution of income: Estimation with limited/incomplete data. *Empirical Economics, 20*, 533–546.

Chotikapanich, D., Griffiths, W. E., & Rao, D. S. P. (2007). Estimating and combining national income distributions using limited data. *Journal of Business and Economic Statistics, 25*, 97–109.

Chotikapanich, D. (Ed.). (2008). *Modeling income distributions and Lorenz curves*. Springer.

Chotikapanich, D., Griffiths, W. E., Rao, D. S. P., & Valencia, V. (2012). Global income distributions and inequality, 1993 and 2000: Incorporating country-level inequality modelled with beta sistributions. *The Review of Economics and Statistics, 94*, 52–73.

Chotikapanich, D., Griffiths, W. E., Karunarathne, W., & Rao, D. S. P. (2013). Calculating poverty measures from the generalized beta income distribution. *Economic Record, 89*(S1), 48–66.

David, H. A., & Nagaraja, H. N. (2003). *Order statistics*. Wiley.

Eckernkemper, T., & Gribisch, B. (2021). Classical and Bayesian inference for income distributions using grouped data. *Oxford Bulletin of Economics and Statistics, 83*, 32–65.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica, 39*, 1037–1039.

Griffiths, W. E., & Hajargasht, G. (2015). On GMM estimation of distributions from grouped data. *Economics Letters, 126*, 122–126.

Hajargasht, G., Griffiths, W. E., Brice, J., Rao, D. S. P., & Chotikapanich, D. (2012). Inference for income distributions using grouped data. *Journal of Business and Economic Statistics, 30*(4), 563–576.

Hajargasht, G., & Griffiths, W. E. (2020). Minimum distance estimation of parametric Lorenz curves based on grouped data. *Econometric Reviews, 39*(4), 344–361.

Hilomi, K., Liu, Q.-F., Nishiyama, Y., & Sueishi, N. (2008). Efficient estimation and model selection for grouped data with local moments. *Journal of the Japan Statistical Society, 38*, 131–143.

Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. Wiley.

Sarabia, J. M., & Jordá, V. (2014). Explicit expressions of the Pietra Index for the generalized function for the size distribution of income. *Physica a: Statistical Mechanics and Its Applications, 416*, 582–595.

Sarabia, J. M., Jordá, V., & Remuzgo, L. (2017). The Theil indices in parametric families of income distributions—A short review. *The Review of Income and Wealth, 63*(4), 867–880.

CHAPTER 6

Empirical Methods for Modelling Economic Insecurity

Nicholas Rohde, Conchita D'Ambrosio, and Barry Watson

Introduction

As a general phenomenon, Economic Insecurity (EI) is seemingly well understood. Most individuals easily recognize the concept and can usually

The unit record data from the HILDA Survey were obtained from the Australian Data Archive, which is hosted by The Australian National University. The HILDA Survey was initiated and is funded by the Australian Government

N. Rohde (✉)
Department of Accounting, Finance and Economics, Griffith Business School, Southport, QLD, Australia
e-mail: n.rohde@griffith.edu.au

C. D'Ambrosio
University of Luxembourg, Esch-sur-Alzette, Luxembourg
e-mail: conchita.dambrosio@uni.lu

B. Watson
University of New Brunswick, Fredericton, NB, Canada
e-mail: bwatson@unb.ca

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_6

relate to instances in their lives where they felt anxious about their short-term economic futures. This sort of anxiety appears to be widespread and occasionally highly intense. For instance, survey data routinely highlight concerns about economic precarity as a major source of personal stress, and a 2015 Pew poll found that more than 90% of respondents preferred stabilizing their economic positions to seeking out upward mobility.[1] These worries can have dramatic implications—for example, financial stress has been cited as reasons for family breakup (Davis & Mantler, 2004), domestic violence (Breiding et al., 2017) and suicide (Burón et al., 2016; Ruhm, 2000).

Nonetheless, EI is rarely discussed in policy circles and remains a relatively new concept in academic research. This is likely due to the concept straddling disciplinary boundaries, sitting somewhere between economics and psychology. But unlike behavioural economics, which considers the effects of psychological processes on economic outcomes, EI does the reverse, examining the impacts of economic factors on psychology. Definitions include Stiglitz et al. (2009), who refer to stresses from "uncertainty about the material conditions that may prevail in the future", while Osberg (1998) defines EI as "the anxiety produced by the lack of economic safety". In later work, Bossert and D'Ambrosio (2013) use "the anxiety produced by the possible exposure to adverse economic events, and by the anticipation of the difficulty to recover from them".

While these definitions are straightforward, producing statistical measures of EI poses some conceptual challenges. Our goal in this chapter is to describe these challenges and provide a review of some of the technical methods developed thus far. We note that several other surveys on EI exist (Osberg, 1998; Richiardi & He, 2019; Rohde & Tang, 2018) and that there are excellent reviews of closely related topics such as vulnerability to poverty (Dercon, 2006). For these reasons, we place our

[1] https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2015/02/americans-financial-security-perceptions-and-reality.

Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views based on the data, however, are those of the author and should not be attributed to the Australian Government, DSS, the Melbourne Institute, the Australian Data Archive or the Australian National University and none of those entities bear any responsibility for the analysis or interpretation of the unit record data from the HILDA Survey provided by the author

emphasis on the more recent and innovative approaches, and those that specifically relate to the anxiety-producing component of economic risk.

Producing good measures of EI is important, as both policy formulation and empirical research are constrained by the quality of the indices. Statistical approaches that fail to properly capture EI may lead to mistargeted policy responses, directing resources towards individuals who may be relatively secure, while missing others in need. And measurement error can bias parameter estimates in econometric models (usually downwards) (Hyslop & Imbens, 2001), meaning that researchers may be ill-equipped to establish relationships between EI and other social phenomena.

While most measures of EI are somewhat ad hoc in nature, they have still been linked to a variety of negative societal outcomes. Applied work in this space has often focussed on implications for health, beginning with Catalano (1991) who documents correlations between EI and a series of mental and physical outcomes (psychological distress, seeking counselling, heart disease). Subsequent research has fleshed these results out, with an increased emphasis on econometric identification. These include studies focusing on obesity, where weight gain represents an evolutionary acquired hard-wired response to stress given overeating historically better equipped individuals to survive periods of food scarcity (Kong et al., 2019; Offer et al., 2010; Smith, 2009; Staudigel, 2016; Watson, 2018; Watson et al., 2020). Other research has focussed on health behaviours (Barnes & Smith, 2011; Case & Deaton, 2015, 2020) or mental health (Kopasker et al., 2018; Rohde et al., 2016; Watson & Osberg, 2017).

EI has also been linked to broader societal problems. The American political scientist Ronald Inglehart (1977) has theorized that economic security is necessary for *postmaterialism*—i.e. a collection of values that emphasize self-expression, freedom of speech, gender equality and environmentalism. In this framework, individuals pursue goals in a hierarchical sequence, and these high-order values cannot be attained until more fundamental needs are met. Unsurprisingly EI is, therefore, related to political behaviour, where it could be expected to encourage support for welfare-state policies (Hacker et al., 2013; Rehm et al., 2012), but also creates right-wing opposition to postmaterialistic policy goals if pursued out of sequence (Bossert et al., 2019; Inglehart & Norris, 2016).

The remainder of the chapter is structured as follows. In section "Measurement Concepts", we outline some of the main concepts important in measuring EI, while section "Subjective Methods" reviews some advantages and disadvantages of using self-assessed empirical approaches. Section “Aggregate Methods” covers early aggregate-level techniques developed in a series of papers by Osberg and Sharpe. section “Axiomatic Methods” and “Micro-econometric Methods” are the most substantial and give an overview of micro-level approaches, while section “Empirical Applications” presents some empirical applications using Australian data. A summary and conclusion are given in section “Conclusion”.

Measurement Concepts

Researchers have developed a number of properties that characterize the nature of EI and can be usefully incorporated into statistical measures (Bossert & D’Ambrosio, 2013; Hacker, 2006; Osberg, 1998). These properties are in some instances abstract, and finding measures that satisfy them is difficult. Nonetheless, the principles provide a conceptual framework for understanding stress-inducing economic risk, and are useful for building an EI measure.

Risk Sensitivity. Economic risks are multifaceted and include (but are not limited to) factors like job loss, income instability, financial problems associated with bereavement or family breakup, and uninsured/unmanageable expenses. Since it is infeasible to quantify all possible risks, a practical approach is to define a *partial* measure of EI that is sensitive to one or more of these potential economic hazards. Further, these hazards should reflect *unwanted* and/or *unforeseen* risks.[2]

Idiosyncrasy. As EI is a psychological phenomenon, personal characteristics linking risk exposure to anxiety are an important part of this process. Some individuals may be unaware of risks they face and are, therefore, not insecure, or may have risk preferences that produce little anxiety. Conversely, others may be deeply risk-averse or prone to experiencing stress and therefore more insecure than their risk profile suggests. An EI measure is *idiosyncratic* if it is sensitive to this heterogeneity.

[2] For example, future joblessness is unlikely to be a source of EI if it reflects an underlying preference or can be planned for well in advance (e.g. a decision to retire).

Prospectivity. Insecurity relates to concerns about the future rather than events experienced in the past. We describe a measure as *prospective* if it relates in some way to an *ex ante*, out-of-sample forecast about economic outcomes, rather than an analysis of historical patterns in data.[3] Note that while *ex ante* measures are explicitly forward-looking, *ex post* or backward-looking measures can still have forward-looking relevance if individuals base their expectations on past experiences.

Relativity. Insecurity is measured in relative terms if risks are defined in proportion to the level of an outcome.[4] Such a property allows insecurity to be independent of scale and defines it in terms of the threat of loss, rather than the threat of deprivation.

Absoluteness. An alternative to the relative characterization of EI is to define it in terms of deprivation. If insecurity is associated with *vulnerability* (the probability of falling into poverty), then measures should be focused on the level of economic outcomes and their absolute changes. Absolute measures of EI should, therefore, decline as individuals become richer, with fears of poverty and deprivation becoming a distant memory.

Subjective Methods

Given the complexities outlined above, perhaps the easiest way to obtain data on individual-level EI is to query respondents about their subjective concerns regarding EI. Since people know their own economic circumstances, have some idea about the types of risks that they may be exposed to, and directly experience their own anxieties, self-assessments allow researchers to access subtleties in information that would otherwise be inaccessible. For example, an individual may have a secure job and a reliable income stream, but may be insecure due to some unobserved reason, such as a health problem that might result in large medical expenses. As

[3] For example, a recently graduated university student may have a historically volatile income stream and hence appear potentially insecure, even if they are in the process of moving into a stable and well-paying job.

[4] A relative measure based upon income volatility would focus on percentage changes to income (e.g. a 20% reduction from one year to the next) rather than unit changes (a $20,000 reduction from one year to another).

analysts will always be unaware of these subtleties, subjective evaluations can allow for both idiosyncratic and prospective measurements.

Nonetheless, there are also a number of drawbacks to using subjective survey responses, which we outline below. In particular, we discuss (i) properties of the data obtained and (ii) heterogeneity in reporting functions, with the corresponding difficulties in performing causal inference using self-assessed variables.

Survey data typically consist of Likert-style questions asking respondents to rate their experiences on a scale (often from 1–5). This technique is routinely employed in psychology and has the benefit of being widely accepted in psychometric research. Two examples related to EI might ask respondents to answer the following:

I worry about making ends meet:

☐	☐	☐	☐	☐
Never (1)	*Occasionally (2)*	*Sometimes (3)*	*Often (4)*	*Always (5)*

I feel insecure about my economic future:

☐	☐	☐	☐	☐
Strongly Disagree (1)	*Disagree (2)*	*Neutral (3)*	*Agree (4)*	*Strongly Agree (5)*

where the responses can be studied individually or aggregated to perform population-level analysis.

Some potential pitfalls to be navigated are outlined. First, note that the data obtained are ordinal, and therefore, researchers might be constrained in the set of statistical techniques available if they wish to preserve this structure. While treating the responses as linear (such that a score of four is double that of two) often yields the same types of results as more complicated ordinal analysis (Norman, 2010), the former practice is often seen as controversial. Second, note the phrasing in the first question explicitly asks about the extensive margin (how often) on EI but not the intensive margin (how much). Therefore, a respondent experiencing an ongoing low-level of EI might give a higher response than someone with intense but intermittent fears. Third, Likert scales are subject to persistent biases in responses such as "left-side preference" and an aversion to giving extreme responses (Chen, 1991). Reversing the order of the questions above will, therefore, produce different responses with subtly higher scoring.

Fourth, and perhaps most importantly, heterogeneous patterns of response can threaten causal inference in research on the social implications of EI. Consider as an example the (stylized) regression model below:

$$v_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + \phi S_i + \varepsilon_i$$

where v_i is an outcome variable of interest (well-being, health, voting behaviour, etc.) and S_i is a survey response such as those outlined above ($\mathbf{x}_i'$ represents a generic set of exogenous controls). The goal of this model is usually to estimate ϕ and interpret it as the impact of S on v. However, since v is almost always either a behavioural variable, or directly influenced by behaviour, unobservable psychological traits will drive both v and S, resulting in biases to ϕ. Without detailed sets of controls or a compelling identification strategy, this form of bias will be difficult to eliminate, making S unsuitable for this form of empirical work.

To better illustrate the roles that risk and response heterogeneity play in generating EI and survey data, we present a conceptual model of the process below. Let R denote the set of all economic risks an individual faces, some of which may be unknown. The function $A_i = f_i(R_i)$ is a mapping from R to A., which is assumed to be a scalar measure of experienced anxiety. Survey methods then capture $S_i = g_i(A_i)$, where $g_i(\cdot)$ is a function coning a subjectively perceived sense of anxiety into a survey response. Combining these means the full function converting risks into survey data is $S_i = g_i(f_i(R_i))$.

We focus on each element in this expression in turn.

- Firstly R_i is the fundamental source of economic insecurity. This is the set of all such economic hazards (e.g. unemployment, unexpected expenses) and will differ across $i = 1, \ldots, n$ due to the varying risk profiles of the population.
- The function $f_i(R_i)$ has a subscript i, indicating that there is heterogeneity (idiosyncrasy) in the ways that risk translates into anxiety. If we assume a functional form such as $A_i = \sum_{r_i \in R_i} \phi_{r_i}$ (i.e. anxiety is the additive sum of all risk perceptions, captured by ϕ_{r_i}), then individuals will differ across ϕ_{r_i}. Notably a risk that an individual is unaware of, or doesn't provoke anxiety, will have $\phi_{r_i} = 0$.
- The function $S_i = g_i(A_i)$ also allows for individual heterogeneity, however, in this instance, the variation in reporting functions is

a disturbance that complicates the study of EI. This function is psychological in nature and represents variations in the ways that perceptions of anxiety are reported. If individual i is stoic and individual j emotional and prone to exaggeration, then $g_i(A_i) < g_j(A_j)$, $A_i = A_j$. Survey responses then desirably capture $f_i(R_i)$ but are contaminated by $g_i(A_i)$. Conversely, econometric estimates (outlined below) only capture a subset of R_i and ignore heterogeneity in $f_i(\cdot)$; however, they are unaffected by $g(\cdot)$.

Aggregate Methods

Much of the seminal work on measuring EI can be traced back to a series of papers by Osberg (1998) and Osberg and Sharpe (2002, 2014), that design and produce a suite of aggregate or country-level risk measures. The idea in these papers is to combine differing macroeconomic indicators of risk exposure as a way of capturing the EI as a latent variable (we also consider latent variable techniques later in the chapter). The techniques involved here are simple, but the results are powerful in their ability to inform economic policy. For example, these aggregate methods are easy to homogenize, and thus are useful for making international or longitudinal comparisons. Statements such as "EI has increased over the last decade" or "Country A is more secure than Country B" are usually the goal of the exercise. Osberg and Sharpe (2014) provide some guidance for producing such indices in rich countries (where high-quality data are available, and risks are more relative in nature) and poor countries (with limited data and absolute risks are more important). We consider each below.

Rich Countries

To measure EI in developed nations the authors employ a "Named Risk" approach, which involves forming a composite index of four separate objective hazards. Citing Article 25 of the UN Declaration of Human Rights (which outlines a series of threats to well-being), they construct risk markers for economic loss due to unemployment, sickness, widowhood, and poverty in old age. Each named risk is itself a combination of sub-indicators, such as health expenditures and insurance rates (illness), protection in light of joblessness (unemployment), and poverty rates, in both head-count and intensity, terms (poverty in old age). They then assume that psychological stress is a function of each input that is constant over countries and over time, such that it can be ignored for the purposes of comparison. The index can then be produced as an average across each dimension, where each dimension is weighted by the fraction of the population affected, that can be scaled to unity.

Poor Countries

The same basic framework can be employed for poorer countries, although allowances are made for (i) differences in data, (ii) changes in the nature and implications of various risks, and (iii) direct deprivation becoming more important as living standards decline. Some data issues are circumvented by merging aggregate variables with results obtained from microdata, such as strata-specific poverty rates or unemployment risks estimated with probit models. Risks are also constructed differently. For example, developing countries may have no social insurance related to unemployment, but rely on informal safety nets via social networks and subsistence agriculture. Poverty depth is more important than for developed countries, and the intensive and extensive margins of poverty are more likely to be correlated with age than in developed nations.

Since aggregate indices in both rich and poor nations represent a combination of (potentially overlapping) risks, the key to their interpretation lies in examining the roles of the variables included. Different sets of named risks will produce different results, and alternative conceptualizations of insecurity (e.g. relative or absolute, distribution-sensitive or population-representative) can have major ramifications. For instance (as we show later), insecurity as characterized as the risk of destitution has almost certainly declined in developed countries over the last few decades, while insecurity associated with unpredictable income volatility has been rising. Ensuring that the choice or indicators match the desired notion of EI is, therefore, fundamental to this approach.

Axiomatic Methods

If researchers want to study the interplay between EI and individual-level characteristics, microeconomic approaches are needed. Here we turn our attention to two methods derived from theory from Bossert and D'Ambrosio (2013) and Bossert et al. (2019). These approaches produce a numerical outcome for each person within a data set summarizing their insecurity at time t. They both also focus on some subset of R_i and use concepts from psychology to construct the mapping onto A_i. The measures share a common structure—each examines short time series of resources (such as income, consumption, or wealth) and produces indices based upon factors such as the order in which observations arrive.

The Bossert et al. (2019) Method

The latter method (Bossert et al., 2019) is both newer and more general, and therefore, we will discuss it first and at greater length. The measure considers a resource stream (income, consumption, wealth, etc.) $y = (y_{-T}, \ldots, y_0) \in \mathbb{R}^T$ where y_0 is the current period and y_{-T}. one T periods in the past. The function $I^T : \mathbb{R}^T \rightarrow \mathbb{R}$ is the insecurity measurewhere higher values indicate a more insecure time profile) and is given meaning by six axioms posited below. The first two axioms are operational in that they define the concept of insecurity captured by I^T. The remaining four are practical and ensure that the index is useful for making interpersonal comparisons.

1. ***Gain-Loss Monotonicity***
   This axiom requires that a gain (loss) in resources from the first period to the second results in ceteris paribus lower (higher) insecurity. This is illustrated with an example drawn from the Bossert et al. (2019) paper below, where two resource streams $y_a(t) = (1, 0, 0, 0)$ and $y_b(t) = (-1, 0, 0, 0)$ are depicted, with the former showing a more insecure state (i.e. $I^4(y_a) > I^4(y_b)$).[5] The individual in the left panel experienced a decline in available resources from $t - 3$ to $t - 2$, while the individual in the right panel reported a corresponding gain. Since gains bring psychological benefits in this model and losses cause psychological costs, y_b is the preferred resource stream evaluated *ex post* at $t = 0$ (Fig. 6.1).
2. ***Proximity Monotonicity***
   This property requires that movements closer to $t = 0$ (the current period) have greater impacts upon the measure. The idea is illustrated with a series of two-period combinations of movements where a gain/loss sequence is followed immediately by a

[5] Note that the intermediate stream $y(t) = (0, 0, 0, 0)$ also needs to yield an intermediate insecurity score.

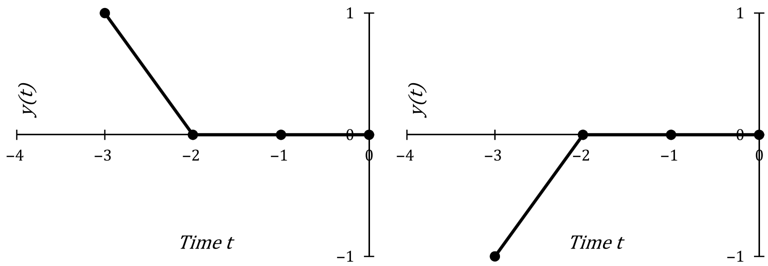

Fig. 6.1 Resource streams $y_a(t) = (1, 0, 0, 0)$ and $y_b(t) = (-1, 0, 0, 0)$

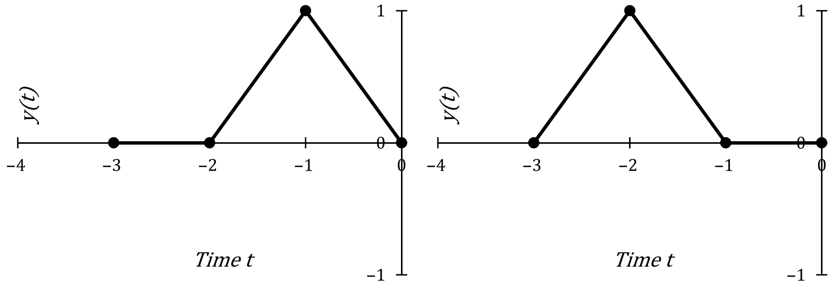

Fig. 6.2 Resource streams $y_c(t) = (0, 0, 1, 0)$ and $y_d(t) = (0, 1, 0, 0)$

reversing loss/gain, capturing the idea that, ceteris paribus, a first-up-then-down move generates insecurity because the individual is discouraged by the immediate loss of a previous gain. Figure 6.2 shows two such resource streams $y_c(t) = (0, 0, 1, 0)$ and $y_d(t) = (0, 1, 0, 0)$, where the former gains in period $t-1$ and loses this gain in t, while the latter occurs one period earlier, in $t-2$ and $t-1$. The dual changes are insecurity-reinforcing in both cases (the losses are more recent than the gain) however in the left panel these occur more recently tohe present. Hence, $I^4(y_c) > I^4(y_d)$.

The reverse case of Fig. 6.2 is presented, where gains are replaced with corresponding losses: $y_e(t) = (0, 0, -1, 0)$ and $y_f(t) = (0, -1, 0, 0)$. Here we observe a first-down-then-up sequence where the individual is encouraged by the immediate recovery from a loss and, therefore, such a move is insecurity-reducing. Hence, in this instance, the dual loss/gain changes *reduce* insecurity, and therefore, proximity to the present is desirable, such that $I^4(y_e) > I^4(y_f)$. Therefore, the ordering of all four streams is $I^4(y_c) > I^4(y_d) > I^4(y_e) > I^4(y_f)$. This highlights a key property of the measure (Fig. 6.3).

3. ***Linear Homogeneity***
   This axiom defines $I^T(y)$ as homogeneous of degree one in y, such that scalar transforms have the property $I^T(\lambda y) = \lambda I^T(y)$. This defines the index in terms of the units used to measure y. Thus, unlike scale-invariant inequality measures such as the Gini coefficient, the Bossert et al. (2019) index requires users to ensure comparability between resource streams, such as correcting for inflation and standardizing using exchange rates.
4. ***Translation Invariance***
   Similar to homogeneity, this property ensures that adding/subtracting a nstant term to each y_t leaves the measure unchanged. This can be stated as $I^T(y_{-T}, \ldots, y_0) = I^T(y_{-T} + \varepsilon, \ldots, y_0 + \varepsilon)$ where $\varepsilon \in \mathbb{R}^1$ is a translation factor. As above, this characterizes insecurity as independent of the level of resources—it is the pattern of change that matters, and hence being well-resourced in an absolute sense offers no protection against insecurity. In conjunction, Axioms 3 and 4 may imply the reverse in practical applications. If intertemporal changes are a function of the levels of $y_{-T}, \ldots, y_0$, then higher values could imply greater absolute losses and gains, and hence higher values of $I^T(y)$.

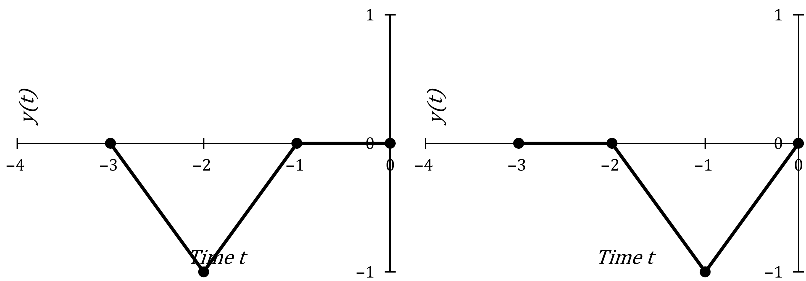

Fig. 6.3 Resource streams $y_e(t) = (0, 0, -1, 0)$ and $y_f(t) = (0, -1, 0, 0)$

5. ***Quasi-Linearity***
Since practical applications of $I^T(y)$ will often involve comparisons where T differs across individuals, a basis for common comparison is required. This axiom requires that if the series $y_{-T}, \ldots, y_0$ is partitioned into sub-streams a and b (e.g. $y_{-T}, \ldots, y_a$ and $y_b, \ldots, y_0$) where then we can write $I^T(y) = f\left(I^{T_a}(y_{-T}, \ldots, y_a); I^{T_b}(y_b, \ldots, y_0)\right)$. By decomposing the measure by sub-stream it can be evaluated over an unbalanced panel without requiring researchers to omit data.
6. ***Stationarity***
In this context, stationarity implies that the measure is independent of the correspondence between $y_{-T}, \ldots, y_0$ and $-T, -T+1, \ldots 0$. Shifting the time period forwards or backwards by a constant integer p leaves $I^T(y)$ unchanged, provided we continue to define the measure at the last observation in the time series.

The authors show that Axioms 1–6 identify the following parametric class of measure:

$$I^T(y) = \ell_0 \left[\sum_{\substack{t \in \{1 \ldots T\} \\ y_{-t} > y_{-(t-1)}}} \delta^{t-1}\left(y_{-t} - y_{-(t-1)}\right) \right] + g_0 \left[\sum_{\substack{t \in \{1 \ldots T\} \\ y_{-t} < y_{-(t-1)}}} \delta^{t-1}\left(y_{-t} - y_{-(t-1)}\right) \right]$$

where parameters ℓ_0 and g_0 weight losses and gains respectively (it is intuitive to set $\ell_0 > g_0$ in order to capture loss-priority) and $\delta < 1$ is a discount factor. According to this index, the least-insecure stream of income is permanently-rising; the most-insecure stream of income is permanently falling. Any constant stream of income produces an insecurity score of zero.

Bossert and D'Ambrosio (2013)

The previous index bears a strong resemblance to the earlier work of Bossert and D'Ambrosio (2013), which is also concerned with interpreting patterns of intertemporal change. In this instance $y = (y_{-T}, \ldots, y_0) \in \mathbb{R}^T$ represents a series of observations on wealth, which is taken to be the full stock of an individual's resources. Again the authors provide an axiomatic structure based upon psychological responses to losses and gains (although we omit it here) and derive the index:

$$V^T(y) = -y_0 + \sum_{\substack{t \in \{1 \ldots T\} \\ y_{-t} > y_{-(t-1)}}} \alpha_{-t}\big(y_{-t} - y_{-(t-1)}\big) + \sum_{\substack{t \in \{1 \ldots T\} \\ y_{-t} > y_{-(t-1)}}} \big(y_{-t} - y_{-(t-1)}\big)$$

In this instance, insecurity is directly related to the current level of resources via $-y_0$ such that *ceteris paribus*, wealthier individuals will be more secure. The functions α_{-t} and β_{-t} are designed to provide weights for losses and gains. These are subjective and can be chosen by the analyst, although Bossert and D'Ambrosio (2013) suggest using $\alpha_{-t} = 1/(2t - 1)$ and $\beta_{-t} = \alpha_{-t}/2$, such that losses are twice as influential as gains (Kahneman & Tversky, 1992). One issue likely to affect implementation of $V^T(y)$ is that y_0 is often some very large number and that the fluctuations in $y_{-T}, \ldots, y_0$ tend to be small. Thus the index is overwhelmingly driven by current wealth. In order to place a greater emphasis on changes rather than levels, researchers may introduce a parameter $\phi \in (01)$ to weight y_0 (D'Ambrosio & Rohde, 2014). Again this is a modelling choice, where higher levels will place more weight on the contributions of past fluctuations, while lower values will emphasize the protective value of a higher wealth buffer.

Micro-Econometric Methods

Hazard Indicators

Most of the applied literature on EI uses micro-econometric techniques to measure exposure to risks, usually related to income or employment. While these indices are often ad hoc, they are also intuitive allowing for clear interpretations, which is a problem for some of the more complex indices. For example, unlike the Bossert et al. (2019) measure, the indices we describe below focus on only one element within R, and assume a constant (and therefore ignorable) relationship back to A.

The most well known of these descriptive techniques comes from Hacker (2006) and Hacker et al. (2011), who focus on downward volatility of income. In a panel of household income data (where y_{it} is now interpreted as current income), they produce the measure:

$$H_{it} = \begin{cases} 1 & \text{if } y_{it} < 0.75 \times y_{it-1} \text{ and } w_{it}, l_{it} \neq 1 \\ 0 & \text{Otherwise} \end{cases}$$

which captures the presence of a decline in income from $t - 1$ to t of 25% or more. As downward volatility in income may not indicate insecurity if the individual is wealthy (denoted with the binary variable w_{it}) or retiring (denoted by l_{it}), these observations are set to zero.[6] Clearly, this is an *ex post* measure as it captures realized volatility, rather than future risk. It is also subject to an arbitrary threshold of 25% and does not allow for graduated levels of insecurity. For these reasons, Hacker (2006) is cautious about interpreting H_{it} as an individual-level measure. Rather, he recommends it be partially aggregated such that levels and trends in insecurity can be estimated for population subgroups based upon identifying characteristics in the microdata.

One way to convert an *ex post* measure such as Hacker's into a prospective or *ex ante* metric is to model the probability of an economically stressful event occurring sometime in the future. If B_{it} represents a binary indicator of a sharp income loss, job loss, major unexpected expense, or other unanticipated shock, then we can use the probability of occurrence, evaluated using lagged covariates as an index. For example, the probit model:

$$P(B_{it} = 1 | \mathbf{x}_{it-1}; \boldsymbol{\beta}) = \phi(\mathbf{x}_{it-1}\boldsymbol{\beta})$$

(where $\phi(\cdot)$ is a Gaussian CDF) will give forward-looking probability estimates of the given hazard materializing. Despite the binary nature of B_{it} the method is quite general and can be used to paint a detailed picture

[6] Other allowances can be made for the presence of safety nets, such as health insurance. See Hacker (2006).

of risk exposure at the individual level. For example, it would be straightforward to define a suite of hazard indicators (smaller and larger income drops, unemployment, falling into poverty, etc.) and assess these risks (i) at a short lag interval such as a year, and over wider timeframes to assess longer term risks. While it may not be feasible to integrate a range of hazard probabilities into a single EI index, a basket of such indicators can provide rich insights into risk exposure without needing to specify the relationship between probabilistic outcomes and EI.

Transitory Variance

A technique for measuring earnings volatility from the labour literature that has received considerable attention is the decomposition of variance into transitory and permanent components (e.g. Gottschalk & Moffit, 2009). While not explicitly defined to measure insecurity (the authors prefer the term "instability" and the measures are not prospective), the concepts are similar enough to be broadly applicable here.[7] Further, a highly attractive feature of these methods is that they blend neatly with the inequality literature, such that inequality due to random fluctuations in income is distinguished from ingrained disparities.

Using the notation above, we wish to write the time series $y_{it-p}, \ldots y_{it}$ as a function of a permanent component $\overline{y}_{it}$ reflecting long-run income or earnings, and a short-run or transitory component $y_{it} - \overline{y}_{it}$. As $\overline{y}_{it}$ itself can adjust (hence the time subscript) it may be estimated using a regression equation, or determined from a moving average based on a window of lagged/leading data. After removing year-specific effects, Gottschalk and Moffit (2009) use the formulae:

$$\text{Transitory Variance} = \sigma_v^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T-1}\sum_{t=1}^{T}\left(y_{it} - \bar{y}_{it}\right)^2\right]$$

$$\text{Permanent Variance} = \sigma_u^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\overline{y} - \overline{\overline{y}}\right)^2 - \sigma_v^2/T$$

[7] Nicholls and Rehm (2014) perform a similar decomposition that they explicitly interpret in terms of EI.

Here σ_v^2 is the average of all individual-level variances, σ_{vi}^2, and σ_u^2 is a measure of permanent income inequality, adjusting for the fact that the full transitory component will not be averaged out over a window of length T. If the data are de-trended at the aggregate level, and if y_{it} is specified in log terms (such that downside movements are more heavily weighted), then σ_{vi}^2 is an indicator of income volatility that will capture some key insecurity concepts, while σ_v^2 represents the corresponding population-level estimate.

Synthetic Indices

Most of the statistical approaches to measuring EI focus on a subset of R (e.g. income volatility as in the Transitory Variance, or unemployment risk) and make the assumption that this dimension broadly captures economic risk as a whole. However, it is possible to use latent variable techniques to extract an estimate of R more broadly. In this instance, EI can be obtained from a series of contaminated indicators (Romaguera-de-la-Cruz, 2019). Suppose we have indicators $q_1 \ldots q_k$ all of which capture some aspect of EI but may also contain unrelated information. The first Principal Component represents a summary of these data, extracting the common element and averaging away the unrelated factors.

To operationalize this technique we z-transform the data $\tilde{q}_j = (q_j - \overline{q}_j)/\sigma_{q_j}$ and construct the linear index

$$\mathbf{w}_1 = \text{argmax}_{\|\mathbf{w}\|=1} \left\{ \sum_{j=1}^{k} \left(w_j \times \tilde{q}_j\right)^2 \right\}$$

where weights $w_1, \ldots . w_k$ are chosen to maximize the variance in $\mathbf{w}_1$ subject to the restriction $\|\mathbf{w}\| = 1$. This optimization process can be undertaken numerically by modern statistical software packages. The advantage of using the fitted value from above is that the interpretation will reflect the underlying components, and hence, it is easy to combine both economic risks, subjective risk assessments and psychological sensitivities in the same measure. A disadvantage is that it can be hard to know whether the extracted component truly has the intended interpretation. For instance, if the covariance in $q_1 \ldots q_k$ more strongly reflect some other common element *besides* EI, then $\sum w_j \times \tilde{q}_j$ will represent this element instead.

A second multidimensional approach comes from Romaguera–de-la-Cruz (2019) and Bucks (2011), which applies the Alkire and Foster (2011) counting technique to EI. In this instance, suppose we have n individuals and $q_1 \ldots q_k$ binary indicators of EI. These may include factors such as (i) being below the poverty rate, (ii) having health insurance, (iii) having ongoing employment, etc. An $n \times k$ exemplar matrix is given below on the left, where the first individual (row) is insecure with respect to indicator q_3 (column 3), the second individual (row) is insecure on q_2 and q_4 (columns 2 and 4), etc. A threshold $h \in \mathbb{N}_+$ is defined such that an individual must be insecure on h dimensions in order to be classified as insecure overall. E.g. if $h = 2$ then all rows with totals less than two have all observations recoded as zeros. In the example this gives the augmented matrix $\boldsymbol{AF}^*$, where the individuals in the top and bottom rows are classed as secure.

$$\boldsymbol{AF} = \begin{bmatrix} 0\,0\,1\,0\,0 \\ 0\,1\,0\,1\,0 \\ 0\,0\,1\,0\,1 \\ 1\,1\,0\,1\,0 \\ 0\,1\,0\,0\,0 \end{bmatrix} \rightarrow \boldsymbol{AF^*}(h_2) = \begin{bmatrix} 0\,0\,0\,0\,0 \\ 0\,1\,0\,1\,0 \\ 0\,0\,1\,0\,1 \\ 1\,1\,0\,1\,0 \\ 0\,0\,0\,0\,0 \end{bmatrix}$$

The Bucks-Romaguera approach (denoted $BR(h)$ here) is to take the average of all elements of AF^*, given by $1/(n \times k)1'\boldsymbol{AF}^*1$. In the above case where $h = 2$ we have a population-level measure $BR(h_2) = 7/25 = 0.28$ while in the below case where $h = 3$, $BR(h_3) = 3/25 = 0.12$.

$$\boldsymbol{BR} = \begin{bmatrix} 0\,0\,1\,0\,0 \\ 0\,1\,0\,1\,0 \\ 0\,0\,1\,0\,1 \\ 1\,1\,0\,1\,0 \\ 0\,1\,0\,0\,0 \end{bmatrix} \rightarrow \boldsymbol{BR^*}(h_3) = \begin{bmatrix} 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 1\,1\,0\,1\,0 \\ 0\,0\,0\,0\,0 \end{bmatrix}$$

Two features of this measure stand out. Firstly, by varying h different levels of intensity are captured. For example, setting $h = 1$ classifies all five individuals as insecure, while using $h = 3$ would only identify the fourth person. By changing the threshold value it is possible to explore ordinal strata of insecurity within a population. Secondly, since the measure is simply the arithmetic mean of the rows and columns of $\boldsymbol{BR}^*$, it is straightforward to write it as a weighted sum of subgroup averages, allowing for intuitive decompositions of the overall index. For example, if the first two rows of $\boldsymbol{BR^*}(h_2)$ are male and the lower three female, then the overall index is $BR(h_2) = (2/5) \times (2/10) + (3/5) \times (5/15)$. Insecurity is then the sum of male and female population weights multiplied by their relative intensities.

Methods Based on Predictive Densities

As most statistical studies of EI focus heavily on income risk, an appealing approach that is implicitly prospective is to use predictive densities to model future outcomes. The key idea here is to use regression models to forecast the full distribution of y for each individual some time (e.g. a year) into the future and then summarize the inherent risk in the distribution. Various forms of this approach exist (e.g. Ligon & Schechter, 2003; Rohde et al., 2020) although we only characterize the method produced by the latter.

Here we specify the model for income based upon lagged covariates predicting the level of income ($\mathbf{x}'_{it-1}$) and its variance ($\mathbf{z}'_{it-1}$)

$$\ln(y_{it}) = \alpha_i + \mathbf{x}'_{it-1}\boldsymbol{\beta} + \varepsilon_{it} \quad \varepsilon_{it} \sim N\left(0, \sigma^2_{it}\right) \quad \sigma^2_{it} = \exp\left(\gamma + \mathbf{z}'_{it-1}\boldsymbol{\theta}\right)$$

Given the normality assumption $\varepsilon_{it} \sim N\left(0, \sigma^2\right)$ this is analogous to fitting the conditional lognormal distribution for each future income with mean $\mu_{it} = \alpha_i + \mathbf{x}'_{it-1}\boldsymbol{\beta}$.

$$f(y_{it}) = \frac{1}{y\sqrt{2\pi \times \exp\left(\gamma + \mathbf{z}'_{it-1}\boldsymbol{\theta}\right)}} \exp\left(-\frac{\ln(y) - \alpha_i - \mathbf{x}'_{it-1}\boldsymbol{\beta}}{2 \times \exp\left(\gamma + \mathbf{z}'_{it-1}\boldsymbol{\theta}\right)}\right)$$

Once modelled in this form, income risk can be summarized in any number of ways. For example, a poverty-based measure may be defined using the probability of falling below some threshold y^*, or the expected shortfall below this point. These two absolute measures are:

$$P\left(y_{it} < y^*\right) = \int_0^{y^*} f\left(y_{it}; \mu_{it}, \sigma^2_{it}\right) dy$$

$$ES(y_{it} < y^*) = \int_0^{y^*} yf\left(y_{it}; \mu_{it}, \sigma_{it}^2\right)dy$$

Conversely, we may wish to summarize risk over the full distribution rather than the left tail. If we define a utility function $U(y)$ such that $U'(y) > 0$ and $U''(y) < 0$, then we can borrow concepts from the inequality literature by comparing utility in this distribution to that of its expected value. If we make the simplifying assumption $U(y) = \ln(y)$ then this ratio leads us to Dalton's (1920) inequality metric, expressed as:

$$D_{it} = 1 - \frac{\mu_{it}}{\mu_{it} + \frac{1}{2}\sigma_{it}^2} \quad D_{it} = 1 - \frac{\alpha_i + \mathbf{x}_{it-1}^{'}\boldsymbol{\beta}}{\alpha_i + \mathbf{x}_{it-1}^{'}\boldsymbol{\beta} + \frac{1}{2}\exp\left(\gamma + \mathbf{z}_{it-1}^{'}\boldsymbol{\theta}\right)}$$

The intuition of this absolute measure is clear. Bound between 0 and 1, the measure captures the percentage of welfare lost to risk. Thus $D_{it} = 0$ when $\sigma_{it}^2 = 0$, and there is no insecurity, as future incomes are known exactly. D_{it} is increasing in σ_{it}^2 and decreasing in μ_{it} where the latter term captures the protective effect from higher incomes. Since the measure considers incomes in their absolutes, we can determine the implicit welfare trade-off between mean and variance. Differentiating with respect to the two arguments gives

$$\frac{\partial D_{it}}{\partial \mu_{it}} = \frac{-2\sigma_{it}^2}{\left(2\mu_{it} + \sigma_{it}^2\right)^2} \quad \frac{\partial D_{it}}{\partial \sigma_{it}^2} = \frac{2\mu_{it}}{\left(2\mu_{it} + \sigma_{it}^2\right)^2}$$

As plausible empirical values are $\mu_{it} = 10$ and $\sigma_{it}^2 = 0.5$ we see that marginal changes to the variance will matter much more for insecurity than marginal changes to the mean. Thus the protective effect of a higher income is relatively low with this measure.

A criticism of Dalton's approach comes from Atkinson (1970) who notes that D is not scale independent and is sensitive to affine transforms on $U(y)$. Both issues can be circumvented if utilities are transformed back into incomes, such that the measure captures the ratio of a certainty-equivalent income to the expected value. Again using $U(y) = \ln(y)$ coupled with a lognormal distribution simplifies this calculation, giving the neat closed-form expression

$$A_{it} = 1 - \exp\left(-\frac{1}{2}\sigma_{it}^2\right) \quad A_{it} = 1 - \exp\left(-\frac{1}{2}\exp\left(\gamma + \mathbf{z}_{it-1}^{'}\boldsymbol{\theta}\right)\right)$$

As above the measure is bounded on the unit interval and captures the proportion of income that could be sacrificed to eliminate economic risk, holding utility constant. Note that the mean $\alpha_i + \mathbf{x}_{it-1}^{'}\boldsymbol{\beta}$ does not appear in these expressions—the index is relative and therefore concerned only with proportional variations in income.

It is also possible to generate absolute indices that capture concepts from prospect theory using this framework. Suppose that individuals adapt rapidly to their current income level y_{it} and their sense of insecurity depends upon the anticipated change for y_{it+1}. Experimental evidence shows that this form of psychological benchmarking is common for smaller risks, and is characterized by (i) L*oss Aversion* and (ii) *Diminishing Sensitivity*. Loss aversion implies that the costs of a negative change are felt more acutely than benefits for a positive change. Diminishing sensitivity describes preferences that are focussed on the presence of a change and are less sensitive to the magnitude (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992).

To incorporate these ideas into an insecurity measure, let us define the change in income $\Delta y_{it} = y_{it+1} - y_{it}$. Since we do not know the future value y_{it+1} we replace it with a predictive density estimated along the lines above. Figure 6.4 highlights these concepts, where *value function V(.)* indicates mood and the horizontal axis the realized change in income. Positive changes are desirable, however, large positive changes bring slightly more well-being than smaller ones. Losses are felt much more acutely (note the steeper slope in the bottom left quadrant); however, increasingly substantial losses again have diminishing negative effects.

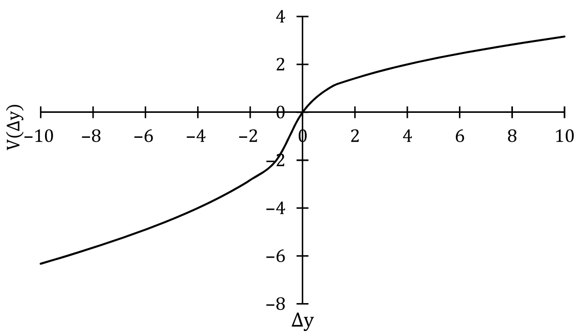

Fig. 6.4 Loss aversion and diminishing sensitivity

By specifying a functional form for $V(.)$ and integrating over the full range of outcomes, we can specify such a measure. For example, the sigmoid function below is useful

$$V(\Delta y) = \begin{cases} \Delta y^{\alpha} \\ -\lambda(-\Delta y)^{\alpha} \end{cases}$$

where $0 < \alpha < 1$ denotes the sensitivity and $\lambda > 1$ loss aversion. Experimental evidence from Kahneman and Tversky (1979) estimates $\alpha = 0.88$ (such that there is little diminishing sensitivity) and $\lambda = 2.25$ (losses are more than twice as strong as gains). We then define a reference dependent measure as the expected change in V as

$$RD_{it}(\Delta y_{it}; \alpha, \lambda) = -\int_{-\infty}^{\infty} V(\Delta y) f(\Delta y_{it}) d\Delta y$$

Empirical Applications

Aggregate Data

In this section, we apply some of the methods outlined in section "Measurement Concepts–Axiomatic Methods". We begin with a small illustrative example of an Osberg-Sharpe style synthetic approach to studying EI at the aggregate level as per section "Measurement Concepts". However, we distinguish our results from those obtained by these authors by drawing a separate set of indicators and make a few methodological adaptions to the construction of the measures. We then look to establish some basic facts about levels and trends in a selected group of high-income countries. In line with Osberg and Sharpe, we measure security rather than insecurity (such that higher values are more desirable), although we note that the interpretations can be reconciled by simply inverting the indices.

Since our goal is to produce longitudinal and cross-national comparisons the demands on our data are relatively heavy, and for this reason, we focus only on a parsimonious set of three indicators (in contrast to Osberg and Sharpe's four "named risks"). And rather than construct specific hazard markers, we use some well-known macroeconomic markers that jointly capture some fundamentals of absolute income risk.

The variables are:

i. Real GDP per capita

ii. Unemployment rates
iii. Government social expenditure as a fraction of GDP.

The idea here is that GDP will capture the relative affluence of the country, which will provide a measure of the distance from (absolute) destitution for the average citizen. The unemployment rate serves as a general proxy for economic risk, noting that other economic hazards such as bankruptcy rates, foreclosures, relative poverty are correlated with this variable. And governmental social expenditure provides a proxy for the size of the welfare state, which serves as an insurance mechanism for individuals in economic distress. Thus, across these indicators we capture a combination of levels, risk exposure and social insurance coverage.

We take OECD data on these indicators and normalize each using z-transforms, which removes scale and variance characteristics specific to each variable. Where appropriate we invert variables such that higher levels indicate greater security. We then produce a convex combination using equal weights (i.e. each series is an arithmetic average of the z-transformed inputs) and plot the results below. Figure 6.5 shows trends in these indices over time for Australia, Canada, Great Britain and the United States, while Fig. 6.5 shows a cross-sectional comparison of the selected OECD countries from 2018.

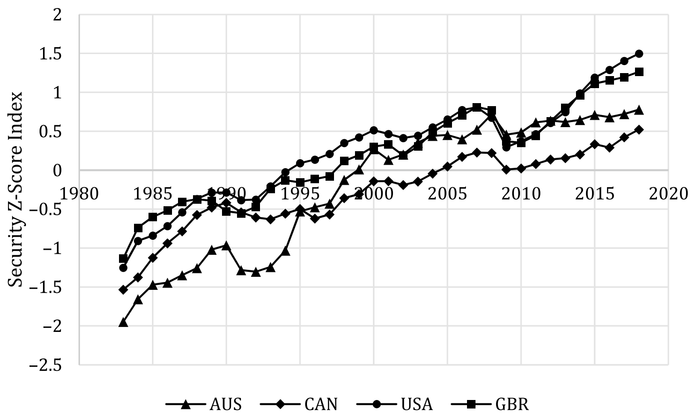

Fig. 6.5 Trends in synthetic (in)security scores 1983–2018

The results from the figure above are fairly intuitive. Since real GDP per capita has been increasing steadily over our 35-year period we expect to see absolute levels of security rise over time. Further, social expenditure as a fraction of GDP has tended to trend upwards over time, and unemployment rates (aside from those observed over the 2008–2012 period) have tended to decline. Thus, aside from (i) the recession in the early 1990s, and (ii) the contraction from 2008 to 2012, we see a pattern of steadily increasing security in all four countries. That the United States is the most secure of the four is surprising, but notably we do not have indicators of within-country distributions of outcomes (such as an absolute poverty rate), which would capture an important element not included here. For example, if insecurity is regarded as a left-tail concept we may expect this ordering to narrow or even reverse due to higher inequalities in the United States. Given economic inequality has been rising in Anglophonic countries since the early 1980s (Alverado et al., 2013) it is also plausible that increasing risks associated with the lower end of the distribution may offset the gains evident here.

Analogous results based upon a cross-section of the same data are given in Fig. 6.6. The snapshot measure depicted uses the same variables employed above, and seems to support anecdotal evidence on the relative security of Nordic and Western European nations. The most secure country is Norway, which has one of the highest levels of GDP per capita in the world, coupled with almost 25% of income allocated to social expenditure. Denmark, Austria, Germany and Belgium reported similarly high values, indicating that economic security cannot be attributed to specifics associated with any single Western European nation. Nonetheless, the European experience is not unique, with Greece remaining highly insecure more than a decade after the financial crisis of 2008, while Spain and Portugal (which were also slow to recover from this recession) were also relatively insecure.

Microdata

In this section, we calculate a selection of individual-level risk/insecurity measures and explore (i) their relationships with each other, and (ii) their correlations with other markers of well-being. We draw our data from the Household Income and Labour Dynamics in Australia (HILDA) survey, which is an approximately representative national panel of around 20,000

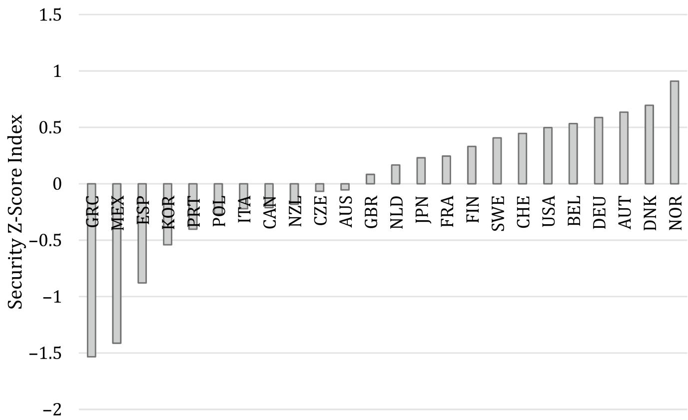

Fig. 6.6 (In)security snapshot—selected OECD countries—2018

individuals followed since 2001. The data set contains a range of variables suitable for measuring economic risk using the methods outlined in section "Aggregate Methods and Axiomatic Methods". We aim to calculate a conceptually diverse set of measures such that differing risk concepts can be assessed. A total of nine measures are used and we briefly describe each below.

1. **Dalton**'s measure of utility lost to income risk is calculated using a one-year prediction window based on real household equivalized income. A pooled regression model is employed to circumvent the estimation of individual-specific effects in a panel data model. Predictions are made using lagged variables capturing educational outcomes, marital status, gender, age, and area of residence.
2. **Atkinson**'s ratio of a certainty-equivalent income to its expected value is used alongside Dalton's measure for household income. Both use the same underlying regression equation as per section "Methods Based on Predictive Densities".
3. A hazard probability of subjectively assessed **Financial Worsening** over one year is estimated. We take a binary indicator of experiencing this hazard and use a probit model as per section "Hazard Indicators" to estimate risk exposure. The same set of covariates is used here as in the Dalton/Atkinson measures.

4. **Probability of Job Loss** in the coming year is a second hazard probability measure estimated with a probit model. This measure is restricted to individuals employed at the time of estimation.
5. A self-assessed measure of **Sufficiency of Emergency Funds** is also drawn. This consists of four ordinal categories referring to the ease at which $3,000 AUD could be raised. While not explicitly a risk marker this measure captures an important dimension of EI through its ability to buffer potential hazards. For simplicity we treat the variable as cardinal.
6. **Self-Assessed Job Security** is captured by a 1–7 scale expressing (dis)agreement with the statement "I have a secure future in my job". Again this measure is restricted to individuals who are employed and we assume cardinality throughout.
7. **Transitory Variance** in household equivalized income is measured using the approach shown in section "Bossert and D'Ambrosio (2013)". We use a rolling average of five-year log incomes to estimate the permanent component. Each individual variance is then calculated relative to this, and the individual variances are averaged to produce the sample-wide estimate.
8. The **Bossert** measure is also calculated using household income streams. We use a lag length of four periods (to preserve a time series long enough to depict trends) and for simplicity we weight equally between gains and losses. Geometric discounting from a base of 0.9 is used to satisfy the proximity property.
9. A **Synthetic Combination** of the previous eight measures is estimated using principal component analysis as per section "Synthetic Indices". A linear prediction is obtained from the first component as a summary of the common element of our basket of measures.

To enable comparison, we perform the same z-transforms to each variable and present some summarizing information below. Figure 6.7 shows the distributions of the transformed measures while Table 6.1 presents the inter-measure correlations. Time trends of each are then depicted below in Figs. 6.8–6.10.

Table 6.1 shows the correlation structure between our nine micro-level measures obtained from HILDA (calculated in the pooled data and assuming cardinality). Aside from the Dalton and Atkinson measures, which are very strongly associated, the empirical links between the

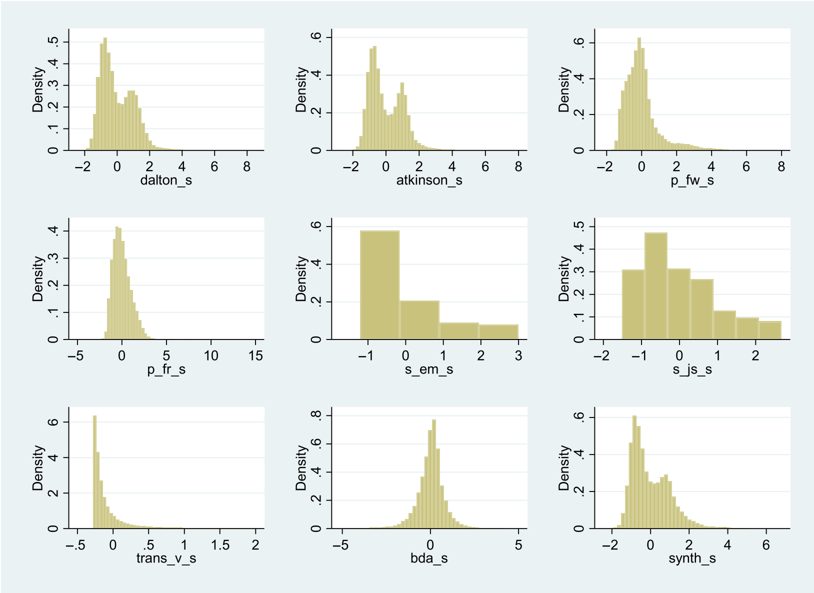

Fig. 6.7 Distributional plots: selected EI indices—HILDA data (*Note* The Figure presents histograms for measures 1–9 depicted across the rows. Measures 1–3 [Dalton, Atkinson, Prob Financial Worsening] in the first row, 3–6 [Prob Job Loss, Emergency Funds Access, Job Security] in the second row, 7–9 the third [Transitory Variance, Bossert Index, 1st PCA]. All indicators are standardized with zero mean and unit variance)

measures are generally small, and sometimes negative. None of the correlations exceed 0.2 and 8/28 pairwise associations are negative (excluding the synthetic summary indicator). This highlights the multidimensionality of EI and serves as something of a cautionary tale. Since there is no strong agreement (and sometimes systematic disagreement) across indices as to who is economically insecure, labelling any single marker as representative of the broader phenomenon seems incorrect. Therefore, it appears that EI may be too complex a concept to boil down to a single representative number. More research is, therefore, required into the function $A_i = f_i(R_i)$, especially in the common instance when A_i is unobserved and R_i only partially observed. Nonetheless, this lack of consistency doesn't leave researchers helpless. Multivariate techniques (such as the

Table 6.1 Correlation matrix: economic risk indices 1–9

Dalton	1.000								
Atkinson	0.988	1.000							
P(F Worse)	−0.039	−0.096	1.000						
P(J Loss)	−0.076	−0.095	0.101	1.000					
Em Funds	−0.032	−0.046	0.114	0.145	1.000				
Job Secure	0.027	0.034	0.024	0.062	0.104	1.000			
Trans V	0.076	0.069	−0.006	0.084	−0.003	0.013	1.000		
Bossert	0.018	0.012	0.034	0.194	0.087	0.056	−0.043	1.000	
1st PC	0.981	0.988	−0.162	−0.189	−0.119	0.031	0.126	−0.026	1.000

Note The table presents Pearson correlation coefficients for all measures performed on the pooled sample. Cardinality is assumed for ordinal measures using data extracted from HILDA

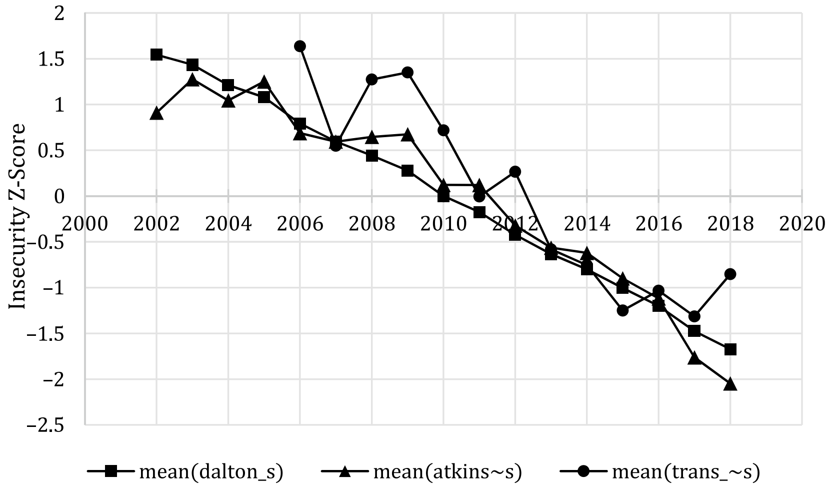

Fig. 6.8 Time trends in micro-data: A

synthetic approaches used above) offer some scope for making more general inferences about EI in these cases, and as empirical results that hold for one indicator may not hold for others, researchers working with several different dimensions can determine where and when inferences about EI are likely to be robust.

Given the small correlations depicted in Table 6.1, is it possible to make generalized claims about the trend of EI in Australia, as we did with the aggregate data? Figs. 6.8–6.10 show annual averages of the normalized indices such that time trends can be assessed. Figure 6.8 shows trends for the three measures based upon income variance (Dalton, Atkinson, Transitory Variance) while Fig. 6.9 shows three measures explicitly linked to psychological perceptions (subjective job security, emergency funds access, BDA). Figure 6.10 shows probabilities of financial worsening, job losses, and the synthetic summary index.

From Panel A it is clear that a general trend can be established for income variance in Australia. Even though the measures are quite conceptually different (Dalton is absolute, Atkinson is relative, and the Transitory Variance is both absolute and *ex post*) there is a steady declining pattern. It, therefore, seems year-to-year fluctuations are declining, and incomes are becoming more predictable over time. Similar patterns can be seen

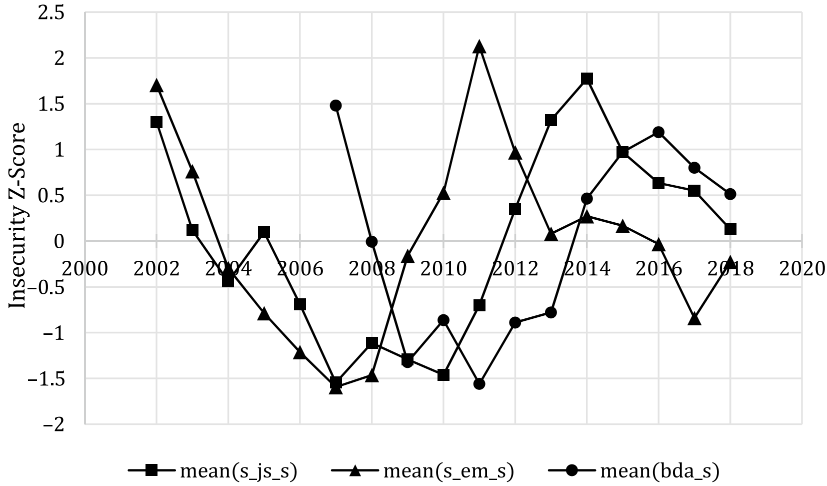

Fig. 6.9 Time trends in micro-data: B

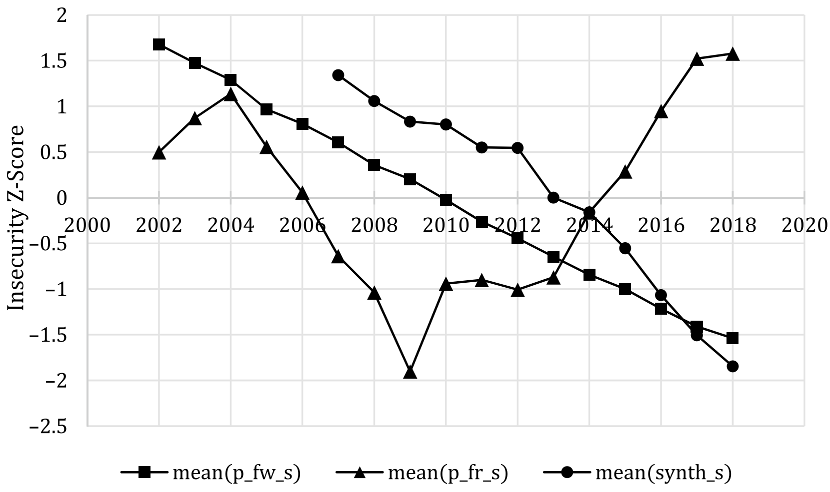

Fig. 6.10 Time trends in micro-data: C

for the probability of a severe financial worsening (Panel C) and the synthetic index (also Panel C). If these measures were the only indicators available, a researcher would be inclined to conclude that EI had fallen unambiguously over recent years.

However, the subjective indicators in Panel B measures (subjective job security, emergency funds access, Bossert index) and the probability of being fired in Panel C tell a more complex story. All four indicators fell steadily until the financial crisis and subsequent global contraction beginning in 2008. In all cases these indices started rising afterwards, either in a sustained way (the probability of being fired) or for several years after (the subjective/Bossert measures). Thus we see a nuanced picture of EI in Australia, where income risk appears to have been steadily falling, but risks related to labour market outcomes, and subjective perceptions of risk are still higher than their long-term averages (zero in the charts) post the 2008 recession.

EI and Social Disadvantage

Despite the difficulties in establishing an unambiguous time trend in EI, there are some empirical regularities that emerge quite strongly in the literature. One such regularity is the association between EI and markers of low socioeconomic status. While we may expect poor people to be more insecure when EI is measured in absolute terms, the result tends to persist with both relative and absolute metrics, as well as subjective and objective measures. To show this, we take the nine indices specified above and average them by socioeconomic subgroup. We stratify individuals into quartiles based upon income levels and produce three educational categories (less than high school; high school; university degree or more) and give the mean outcomes in Table 6.2.

The associations between income and all nine measures are evident in Table 6.2. Our poorest income group (Quartile 1) had above average (positive) scores across the full range of indices, and in all instances had higher outcomes than the corresponding figures for Quartile 2. Scores in Quartile 3 were almost uniformly higher than Quartile 2, and all values aside from the synthetic index were negative for Quartile 4. Almost half of the individuals in Q1 are below Australia's relative poverty line (11.2% of our data fall below half the median income) the poor and near-poor are especially disadvantaged when it comes to EI. A similar result can be obtained with respect to education, with individuals with university

Table 6.2 Socioeconomic status and insecurity indices

	Income				*Education*		
	INC Q1	*INC Q2*	*INC Q3*	*INC Q4*	*< HS*	*HS*	*> HS*
Dalton	0.2644	0.0111	−0.1157	−0.1598	0.3482	0.0171	−0.2775
Atkinson	0.1232	−0.0240	−0.0678	−0.0315	0.2946	0.0027	−0.2153
P(F Worse)	0.7140	0.0901	−0.2171	−0.5865	0.1481	0.0198	−0.1389
P(J Loss)	0.3054	0.0591	−0.0567	−0.3075	−0.0553	0.1775	−0.2550
Em Funds	0.3027	0.0961	−0.0708	−0.3276	0.3257	0.0576	−0.3287
Job Secure	0.1064	−0.0008	−0.0341	−0.0715	0.0743	0.0065	−0.0640
Trans V	0.2470	−0.0423	−0.0640	−0.0254	0.0069	0.0083	−0.0172
Bossert	0.6924	0.3860	0.1542	−0.5381	0.1193	0.0515	−0.1276
1st PC	0.1094	−0.0957	−0.0736	0.0889	0.1980	−0.0104	−0.0796

Note The table gives averages of all nine indices averaged over sample subgroups defined on the basis of socioeconomic status. Q1 (1st column) refers to observations in the lowest income quartile and Q4 (4th column) refers to the highest quartile. Similarly, educational groups (less than high school; high school; tertiary) appear in columns 7–7. All indices are normalized such that a value of zero represents the sample average

educations the most secure on every indicator while eight from nine measures are highest for persons with less than high school education.

Regression Models

Lastly, we replicate some standard results showing that EI is predictive of diminished well-being in observational data. The results are descriptive rather than causal as we do not employ a specific identification strategy (besides the use of a standard battery of controls), however, causation has been established by a variety of authors in other contexts (e.g. Kong et al., 2019; Staudigel, 2016). The purpose here instead is to perform a validation exercise—if our measures are genuinely capturing an important well-being concept, then we expect empirical links to exist between EI and other markers of welfare. Below we run OLS regressions of the form outlined at the start of the chapter:

$$v_i = \alpha + \lambda_t + \mathbf{x}_i^{'}\boldsymbol{\beta} + \phi EI_i + \varepsilon_i$$

As above, α is an intercept, λ_t a time-effect (allowing health and satisfaction to evolve over time), $\mathbf{x}_i^{'}$ a vector of controls and EI_i an insecurity measure. Parameter ϕ, therefore, captures the association between the outcome variable v and EI conditioned on the other covariates.

The results are reported in Tables 6.3 and 6.4, where the former uses 0–10 life satisfaction outcomes and the latter (0–100) SF-36 mental health scores. In both instances, higher values on the LHS indicate greater well-being, and therefore, we expect to see negative coefficients on the EI metrics in the first nine rows. Again we use the z-transformed measures such that each indicates the effect of a standard deviation shift, which allows for comparisons to be performed across the columns.

The parameter estimates for life satisfaction (Table 6.3) show that all measures of EI are negatively correlated with the dependent variable, after conditioning on factors such as current income and educational attainment. These range between −0.265 (Self-Assessed Job Security) and −0.011 (Transitory Variance) and are significant at standard levels. A standard deviation shock to these risk markers, therefore, predicts a decline in life satisfaction of about 0–0.26 points, with the summary index having an effect size of −0.187 units per SD. Since current socioeconomic status is controlled in the regressions, this suggests that our parameter estimates are capturing a risk effect rather than economic deprivation, which is the intent of the measures.[8]

There are some attractive similarities between these estimates and those obtained for the SF-36 mental health aggregates in Table 6.4. Again the parameter estimates are always negative, with the largest effect size on Self-Assessed Job Security (−3.225) and the lowest on the Transitory Variance (−0.730). The latter is not intended as an ex ante metric; this may explain the relatively low conditional associations. The results also show greater magnitudes for the subjectively assessed measures, although as we point out these are less likely to be causal. As the units of measurement for the SF-36 are different to those used for life satisfaction, we cannot directly compare the results across the two tables, but again by standardizing (this time by using the ratios of the standard deviations of the two dependent variables), we see that the estimates on average are around 50% higher for life satisfaction than for mental health. Economic insecurity seems, therefore, to be a little more closely linked with unhappiness than it does poor mental health.

[8] Note that we exclude current income as a control in the Bossert measure as this appears directly in the EI measure.

Table 6.3 Life satisfaction and economic insecurity indices

	M(1)	*M(2)*	*M(3)*	*M(4)*	*M(5)*	*M(6)*	*M(7)*	*M(8)*	*M(9)*
Dalton	−0.235***								
Atkinson		−0.228***							
P(F Worse)			−0.130***						
P(J Loss)				−0.205***					
Em Funds					−0.190***				
Job Secure						−0.265***			
Trans V							−0.011**		
Bossert								−0.037***	
1st PC									−0.187***
Log Inc	0.216***	0.221***	0.078***	0.104***	0.155***	0.187***	0.232***		0.249***
Female	0.007	0.006	0.050***	−0.174***	0.064***	0.023***	0.027***	0.038***	0.043***
Age	−0.142***	−0.141***	−0.053***	−0.072***	−0.080***	−0.076***	−0.066***	−0.066***	−0.124***
Age Sq	0.002***	0.002***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
Post Grad	−0.087***	−0.050***	−0.127***	−0.135***	−0.155***	−0.069***	−0.078***	0.040*	−0.007
Degree	−0.127***	−0.103***	−0.027*	−0.093***	−0.108***	−0.069***	−0.034**	0.009	−0.099***
Tertiary	−0.073***	−0.068***	−0.082***	−0.052***	−0.141***	−0.088***	−0.043***	−0.028*	−0.070***
H School	−0.039***	−0.039***	−0.079***	−0.042***	−0.097***	−0.074***	−0.037***	−0.025	−0.013
W2	0.039	0.047*	0.053**	0.065***	0.047*	0.033			
W3	−0.019	−0.002	0.016	0.036	0.001	−0.014			
W4	−0.126***	−0.101***	−0.077***	−0.048**	−0.094***	−0.099***			
W5	−0.162***	−0.130***	−0.092***	−0.056**	−0.110***	−0.127***			
W6	−0.166***	−0.125***	−0.074***	−0.032	−0.096***	−0.124***	0.005		
W7	−0.212***	−0.164***	−0.104***	−0.057**	−0.124***	−0.150***	−0.029	−0.003	−0.027
W8	−0.204***	−0.149***	−0.079***	−0.025	−0.085***	−0.129***	−0.016	0.032	−0.021
W9	−0.258***	−0.195***	−0.111***	−0.047**	−0.115***	−0.167***	−0.039*	−0.015	−0.088***
W10	−0.242***	−0.171***	−0.081***	−0.007	−0.068***	−0.128***	−0.048**	0.011	−0.087***

	M(1)	*M(2)*	*M(3)*	*M(4)*	*M(5)*	*M(6)*	*M(7)*	*M(8)*	*M(9)*
W11	−0.255***	−0.177***	−0.075***	0.005	−0.072***	−0.114***	−0.031	0.02	−0.099***
W12	−0.312***	−0.226***	−0.113***	−0.024	−0.118***	−0.141***	−0.066***	0.001	−0.140***
W13	−0.298***	−0.205***	−0.085***	0.012	−0.089***	−0.109***	−0.035	0.019	−0.142***
W14	−0.295***	−0.196***	−0.064***	0.041*	−0.071***	−0.103***	−0.032	0.061**	−0.118***
W15	−0.336***	−0.230***	−0.090***	0.023	−0.101***	−0.136***	−0.060***	0.061**	−0.132***
W16	−0.389***	−0.276***	−0.125***	−0.004	−0.144***	−0.175***	−0.087***	0.037	−0.173***
Constant	8.344***	8.187***	8.077***	8.406***	7.871***	7.410***	6.630***	9.216***	7.558***
R^2	0.039	0.039	0.032	0.036	0.047	0.071	0.025	0.014	0.029
F	186.16	184.696	135.246	164.235	201.547	321.957	103.711	29.324	54.094
N	121,029	121,029	121,029	121,029	121,029	121,029	84,938	34,233	34,233

Note The table presents results of pooled OLS regressions of 0–10 life satisfaction scores on normalized insecurity indices1-9 and a collection of socioeconomic controls. Robust covariance is used for inference. Symbols *,** and *** denote significance at 10*, 5* and 1* respectively. W2–W16 reflect year-specific fixed-effects

Table 6.4 Mental health and economic insecurity indices

	M(1)	*M(2)*	*M(3)*	*M(4)*	*M(5)*	*M(6)*	*M(7)*	*M(8)*	*M(9)*
Dalton	−1.774***								
Atkinson		−1.742***							
P(F Worse)			−0.894***						
P(J Loss)				−1.473***					
Em Funds					−2.804***				
Job Secure						−3.225***			
Trans V							−0.073		
Bossert								−0.311***	
1st PC									−1.604***
Log Inc	2.294***	2.333***	1.357***	1.490***	1.248***	1.848***	2.330***		2.314***
Female	−2.456***	−2.461***	−2.135***	−3.745***	−1.829***	−2.389***	−2.045***	−1.341***	−1.297***
Age	−0.736***	−0.734***	−0.076***	−0.207***	−0.328***	−0.256***	−0.213***	−0.265***	−0.757***
Age Sq	0.011***	0.011***	0.003***	0.003***	0.005***	0.004***	0.004***	0.004***	0.011***
Post Grad	−0.236	0.052	−0.508***	−0.578***	−1.220***	−0.002	−0.464**	0.147	−0.323
Degree	−0.470***	−0.293*	0.274	−0.198	−0.772***	−0.133	−0.116	−0.074	−1.044***
Tertiary	0.123	0.156	0.061	0.268**	−0.874***	−0.055	0.038	−0.237	−0.640***
H School	0.588***	0.586***	0.292**	0.556***	−0.074	0.278**	0.449***	0.263	0.355
W2	−0.293	−0.23	−0.188	−0.097	−0.249	−0.402			
W3	−0.691**	−0.568**	−0.436	−0.289	−0.571**	−0.743***			
W4	−1.240***	−1.057***	−0.885***	−0.671**	−1.011***	−1.071***			
W5	−1.246***	−1.004***	−0.732***	−0.468*	−0.827***	−1.043***			
W6	−1.582***	−1.279***	−0.912***	−0.597**	−1.007***	−1.357***	−0.254		
W7	−1.984***	−1.625***	−1.193***	−0.842***	−1.195***	−1.560***	−0.464*	−0.3	−0.524
W8	−1.825***	−1.410***	−0.908***	−0.503*	−0.663**	−1.295***	−0.275	0.252	−0.234
W9	−2.489***	−2.014***	−1.413***	−0.939***	−1.087***	−1.843***	−0.811***	−0.235	−0.900**
W10	−3.060***	−2.528***	−1.879***	−1.332***	−1.278***	−2.179***	−1.305***	−0.516	−1.404***

	M(1)	*M(2)*	*M(3)*	*M(4)*	*M(5)*	*M(6)*	*M(7)*	*M(8)*	*M(9)*
W11	−3.081***	−2.495***	−1.760***	−1.164***	−1.246***	−1.922***	−1.231***	−0.161	−1.244***
W12	−3.409***	−2.768***	−1.953***	−1.293***	−1.519***	−1.947***	−1.363***	−0.339	−1.617***
W13	−4.053***	−3.359***	−2.493***	−1.771***	−2.013***	−2.428***	−2.032***	−0.794**	−2.258***
W14	−4.587***	−3.842***	−2.890***	−2.116***	−2.415***	−2.972***	−2.390***	−1.276***	−2.892***
W15	−4.950***	−4.152***	−3.144***	−2.306***	−2.685***	−3.291***	−2.626***	−1.088***	−2.837***
W16	−5.439***	−4.592***	−3.504***	−2.608***	−3.116***	−3.670***	−3.115***	−1.758***	−3.655***
Constant	64.550***	63.509***	61.663***	64.447***	68.861***	60.415***	53.445***	80.465***	64.287***
R^2	0.0300	0.03	0.027	0.029	0.053	0.068	0.025	0.017	0.024
F	159.515	159.133	142.84	153.865	250.446	342.119	114.508	36.853	46.416
N	121,029	121,029	121,029	121,029	121,029	121,029	84,938	34,233	34,233

Note The table presents results of pooled OLS regressions of 0–100 SF-36 mental health scores on normalized insecurity indices1-9 and a collection of socioeconomic controls. Robust covariance is used for inference. Symbols *,** and *** denote significance at 10*, 5* and 1* respectively. W2-W16 reflect year-specific fixed-effects

Conclusion

The chapter has presented a conceptual overview of economic insecurity and discussed a variety of methods used for measurement. Our review is far from exhaustive, however, it serves as an introduction to the key themes that populate this literature. In particular we covered self-assessed risk measures, axiomatic methods, multivariate indicators, and a range of income risk indices such as hazard probabilities and those based on the predictive variance. We note that none of these approaches fully capture EI, and that all methods have some advantages and disadvantages.

We see future work on economic insecurity proceeding in two different directions. Firstly, given that the multifaceted nature of EI is so important, finding ways to simultaneously study multiple risks seems an obvious way forward. While multivariate techniques such as PCA, or the counting technique from Alkire and Foster (2011), Bucks (2011) and Romaguera-de-la-Cruz (2019) offer promise, other methods such as the "dashboard" approach advocated by Ravallion (2011) for poverty analysis are also worth pursuing. Here the analyst does not try to measure EI per se, but rather focuses on a set of important risk indicators without trying to combine them. This approach neglects the covariance structure of the indicators, but as we have seen the inter-measure correlations tend to be quite small.

Other developments in measurement will ideally produce a wider variety of idiosyncratic measures. The only indices we covered here that have this property is the self-assessed measures from section "Subjective Methods", which we argued were often unsuitable for applied work in other ways. One promising option for incorporating some subjective element into an EI index is to take psychometric data on anxiety and regress this against a basket of risk indicators and controls. The variation in anxiety explained by the risk comes fairly close to the conceptual definition of EI, and heterogeneity in the function $A_i = f_i(R_i)$ can be partially incorporated by running separate regressions on stratified data.

Secondly, we anticipate better measures to produce richer empirical work. For example, it is widely believed that EI is a contributing factor behind the opioid epidemic in the United States, and some of the social schisms and populist political movements that have emerged in developed countries over the last few years. Determining what roles (if any) EI plays will require careful applied research, but progress in this space offers scope for addressing a set of damaging social problems.

Acknowledgements We thank Prasada Rao for his ongoing support, and Alannah Pocknall for research assistance

References

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics, 95*, 476–487.

Alvaredo, F., Atkinson, A., Piketty, T., & Saez, E. (2013). The top 1% in international and historical perspective. *Journal of Economic Perspectives, 27*, 3–20.

Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory, 2*, 244–263.

Barnes, M., & Smith, T. (2011). Tobacco use as response to economic insecurity: Evidence from the national longitudinal survey of youth. *The B.E. Journal of Economic Analysis & Policy, 9*(1).

Bossert, W., Clark, A., D'Ambrosio, C., & Lepinteur, A. (2019). *Economic insecurity and the rise of the right* (CEP Discussion Paper No. 1659).

Bossert, W., & D'Ambrosio, C. (2013). Measuring economic insecurity. *International Economic Review, 54*, 1017–1030.

Breiding, M. J., Basile, K., Klevens, J., & Smith, S. (2017). Economic insecurity and intimate partner and sexual violence victimization. *American Journal of Preventative Medicine, 53*, 457–464.

Bucks, B. (2011). *Economic vulnerability in the United States: Measurement and trends*. Paper Prepared for the IARIW-OECD Conference on Economic Insecurity.

Burón, P., Jimenez-Trevino, L., Saiz, P., García-Portilla, M., Corcoran, P., Carli, V., Fekete, S., Hadlaczy, G., Hegerl, U., Michel, K., Sarchiapone, M., Wasserman, D., Schmidtke, A., & Bobes, J. (2016). Reasons for attempted suicide in Europe: Prevalence, associated factors, and risk of repetition. *Archives of Suicide Research, 20*, 45–58.

Case, A., & Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-hispanic Americans in the 21st century. *Proceedings of the National Academy of Science, 112*(49), 15078–15083.

Case, A., & Deaton, A. (2020). *Deaths of despair and the future of capitalism*. Princeton University Press.

Catalano, R. (1991). The health effects of economic insecurity. *American Journal of Public Health, 81*, 1148–1152.

Chen, J. (1991). Response-order effects in likert-type scales. *Educational and Psychological Measurement, 51*, 531–540.

D'Ambrosio, C., & Rohde, N. (2014). The distribution of economic insecurity: Italy and the U. S. over the great recession. *Review of Income and Wealth, 60*, S33–S52.
Dalton, H. (1920). The measurement of the inequality of incomes. *Economic Journal, 30*, 348–461.
Davis, G., & Mantler, J. (2004). *The consequences of financial stress for individuals, families and society*. Centre for Research on Stress, Coping and Well-being.
Dercon, S. (2006). Vulnerability: A micro perspective. In F. Bourguignon, B. Pleskovic, & J. van der Gaag (Eds.), *Securing development in an unstable world*. World Bank publications.
Gottschalk, P., & Moffitt, R. (2009). Household risks: The rising instability of U.S. earnings. *Journal of Economic Perspectives, 23*, 3–24.
Hacker, J. (2006). *The great risk shift*. Oxford University Press.
Hacker, J., Huber, G., Rehm, P., Schlesinger, M., & Valletta, R. (2011). *Economic security at risk: Findings from the economic security index*. Rockefeller Foundation.
Hacker, J., Rehm, P., & Schlesinger, M. (2013). The insecure American: Economic experiences, financial worries, and policy attitudes. *Perspectives on Politics, 11*, 23–49.
Hyslop, D., & Imbens, G. (2001). Bias from classical and other forms of measurement error. *Journal of Business and Economic Statistics, 19*, 475–481.
Inglehart, R. (1977). *The silent revolution: Changing values and political styles among western publics*. Princeton University Press.
Inglehart, R., & Norris, P. (2016). *Trump, Brexit, and the rise of populism: Economic have-nots and cultural backlash* (HKS Working Paper No. RWP16–026).
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.
Kong, N., Osberg, L., & Zhou, W. (2019). The shattered "Iron Rice Bowl": Intergenerational effects of Chinese state-owned enterprise reform. *Journal of Health Economics*. https://doi.org/10.1016/j.jhealeco.2019.06.007
Kopasker, D., Montagna, C., & Bender, K. (2018). Economic insecurity: A socioeconomic determinant of mental health. *SSM—Population Health, 6*, 184–194.
Ligon, E., & Schechter, L. (2003). Measuring vulnerability. *Economic Journal, 113*(486), C95–C102.
Nichols, A., & Rehm, P. (2014). Income risk in 30 countries. *Review of Income and Wealth, 60*, S98–116.
Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education Theory and Practice, 15*, 625–632.

Offer, A., Pechey, R., & Ulijasazek, S. (2010). Obesity under affluence varies by welfare regimes: The effect of fast food, insecurity, and inequality. *Economics and Human Biology, 8*, 297–308.

Osberg, L. (1998). *Economic insecurity*. Social Policy Research Centre, UNSW.

Osberg, L., & Sharpe, A. (2002). An index of economic well–being for selected OECD countries. *Review of Income and Wealth, 48*, 291–316.

Osberg, L., & Sharpe, A. (2014). Measuring economic insecurity in rich and poor nations. *Review of Income and Wealth, 60*, 53–76.

Ravallion, M. (2011). On multidimensional indices of poverty. *Journal of Economic Inequality, 9*, 235–248.

Richiardi, M., & He, Z. (2019). *Measuring economic insecurity: A review of the literature*. CeMPA WP 1/20.

Rehm, P., Hacker, J., & Schlesinger, M. (2012). *Insecure alliances: Risk, inequality, and support for the welfare state. American Political Science Review, 106*, 386–406.

Rohde, N., & Tang, K. K. (2018). Economic insecurity: Theoretical approaches. In C. D'Ambrosio (Ed.), *Handbook of research on economic and social well-being* (pp. S300–S315). Edward Elgar.

Rohde, N., Tang, K., D'Ambrosio, C., Osberg, L., & Rao, P. (2020). Welfare-based income insecurity in the US and Germany: Evidence from harmonized panel data. *Journal of Economic Behavior & Organization, 176*, 226–243.

Rohde, N., Tang, K. K., Osberg, L., & Rao, P. (2016). The effect of economic insecurity on mental health: Recent evidence from Australian panel data. *Social Science & Medicine, 151*, 250–258.

Romaguera-de-la-Cruz, M. (2019). Measuring economic insecurity using a counting approach. an application to three EU countries. *Review of Income and Wealth*. https://doi.org/10.1111/roiw.12428

Ruhm, C. (2000). Are recessions good for your health? *Quarterly Journal of Economics, 115*, 617–650.

Smith, T. (2009). Reconciling psychology with economics: Obesity, behavioral biology, and rational overeating. *Journal of Bioeconomics, 11*, 249–282.

Staudigel, M. (2016). A soft pillow for hard times? Economic insecurity, food intake and body weight in Russia. *Journal of Health Economics, 50*, 198–212.

Stiglitz, J., Sen, A., & Fitoussi, J. (2009). *Report by the commission on the measurement of economic performance and social progress*. http://www.stiglitzsen-fitoussi.fr/en/index.htm

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

Watson, B. (2018). Does economic insecurity cause weight gain among Canadian labor force participants? *Review of Income and Wealth, 64*, 406–427.

Watson, B., & Osberg, L. (2017). Healing and/or breaking? The mental health implications of repeated economic insecurity. *Social Science and Medicine, 188*, 119–127.

Watson, B., Daley, A., Rohde, N., & Osberg, L. (2020). Blown off-course? Weight gain among the economically insecure during the great recession. *Journal of Economic Psychology*, In Press.

CHAPTER 7

Measuring Inequality in Health

Bénédicte Apouey and Jacques Silber

Introduction

A substantial literature in economics and social sciences shows that a higher socioeconomic status (SES) is associated with better health outcomes (Apouey, 2015; Deaton, 2013; Kivimäki et al., 2020; O'Donnell et al., 2015). Differences in morbidity, mortality, and health care use according to social position are systematic and pervasive over time and space. There is some evidence that crises, such as the Great Recession and the COVID-19 pandemic, exacerbate existing social health inequalities over the world (Marmot & Allen, 2020). Following Deaton (2013),

---

We dedicate this paper to the memory of Adam Wagstaff who was a pioneer in the study of health inequality and who passed away prematurely in May 2020.

B. Apouey
Paris School of Economics—CNRS, Paris, France
e-mail: benedicte.apouey@psemail.eu

J. Silber (✉)
Bar-Ilan University, Ramat Gan, Israel
e-mail: jsilber_2000@yahoo.com

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*, https://doi.org/10.1007/978-981-19-2023-3_7

health inequalities do matter, and there are cases in which they may be considered as unjust: "*I treat health inequalities as important to the extent that they involve inequalities in overall wellbeing, and treat health inequalities as unjust when they are not compensated for by other components of wellbeing, and when they are remediable, but not remediated*." Because the correlation between the two variables of interest (socioeconomic status and health) may not only capture a unidirectional causal effect from one variable to the other one, but may also involve reverse causation and more complex pathways between variables, the causes of social health inequalities are not yet perfectly understood.

In this chapter, we present the literature on the *measurement* of overall and social health inequalities. Measuring inequalities and understanding their origins are a prerequisite for implementing an efficient policy aiming at reducing them. The chapter carefully describes the axiomatic properties of a number of inequality indicators, highlighting the normative hypotheses underlying them. We distinguish between cardinal and ordinal health variables, and between the univariate and bivariate approaches.

At the individual level, cardinal health variables include the McMaster Health Utility Index (HUI, which captures individual functional health and varies between 0 and 1), body length measures (e.g., child height-for-age percentile score), or the number of chronic conditions, for instance.[1] Importantly, cardinal data are not always available. In particular, the general health status is generally captured using an ordinal variable, derived from the following question: "How is your health in general?", with the following response categories: "(1) very good," "(2) good," "(3) fair," "(4) bad," and "(5) very bad."[2] Self-assessed health is a valuable health measure because it provides a summary of individual health status, including physical and mental components. Moreover, it is a good predictor of future mortality, even when controlling for individual health indicators and sociodemographic characteristics (Idler & Benyamini, 1997). However, a potential limitation is that self-assessed health may

[1] According to Erreygers and Van Ourti (2011a), there are different types of measurement scales (ordinal, cardinal, ratio-scale, and fixed). Authors also distinguish between bounded and unbounded variables. In this perspective, while body temperature or the HUI are measured on a cardinal scale, health care expenditures and body length are ratio-scale variables, and the number of chronic conditions or of doctor visits are measured on a fixed scale.

[2] This wording of the question is recommended by the WHO Regional Office for Europe and provides a basis for comparisons of self-assessed health across countries.

suffer from reporting heterogeneity: in other words, individuals with the same level of true (unobserved) health, but with different sociodemographic characteristics or past illnesses, may systematically answer the self-assessed health question in different ways.

Conventional tools to measure inequality (such as the Gini coefficient), which are mean-based, are suited for cardinal health variables but cannot be directly used for ordinal outcomes. In the measurement of health inequality for ordinal outcomes, the median category, rather than the mean health status, plays generally an important role. The theoretical literature on the measurement of inequality for cardinal variables is a well-established line of research, while articles on the measurement using ordinal outcomes are more recent. Note that a strand of the literature converts ordinal health variables to cardinal ones, to allow the use of traditional tools which are suited for cardinal outcomes. Several techniques are available to convert data. For instance, an ordinal health variable may be replaced with a predicted health score derived from a regression. As an example, Van Doorslaer and Jones (2003) transform the ordinal self-assessed health variable into a continuous variable by regressing it on a number of determinants, using an ordered probit model, computing the prediction, and rescaling this prediction. They also show that an interval regression (using the HUI to set the bounds of self-assessed health) may also be used to construct a cardinal health score; in that case, no rescaling is necessary. An advantage of this conversion is that it partially eliminates heterogeneity in responses to the self-assessed health question. Whatever the transformation method selected, the quality of the conversion is always questionable. On a related matter, conversions do not respect the ordinal nature of the original data. In this chapter, we distinguish between inequality measures for cardinal and ordinal health outcomes. Moreover, we draw a distinction between individual-level and grouped health data.

The univariate and bivariate approaches to inequality measurement represent two different strands in the literature (Wagstaff & Van Doorslaer, 2004). While the univariate approach considers health variations within the population without any reference to the distribution of socioeconomic status ("overall health inequality"), the bivariate perspective assesses variations in health according to socioeconomic status ("social" or "socioeconomic health inequality"). Economists generally focus on the bivariate approach and extensively rely on one specific measure, namely the concentration index, to capture this type of inequality. This is due to the fact that the concentration index has a number of interesting features: in particular, it can be decomposed into factors to unravel the causes of inequalities. However, this chapter presents a number of alternative indices in addition to the concentration index.

The recent literature provides evidence on the existence of (social or overall) health inequality for many countries, including low-income countries. Moreover, a number of international comparisons of health inequalities have already been made (Le Grand, 1987; Van Malderen et al., 2019). In this chapter, we present some of these empirical studies, highlighting the choice of indices and findings. These articles employ data on African, Asian, and European countries, which differ in terms of key health indicators (i.e., mortality rate, average health status, etc.), health care systems and access to health care, and levels and trajectories of socioeconomic inequality (e.g., income inequality).

The remainder of this chapter is structured as follows. Section "Cardinal Variables and the Measurement of Inequality in Health" reviews the literature on the measurement of overall and social health inequalities for cardinal health outcomes. Section "Measuring Health Inequality and Polarization with Ordinal Variables" focuses on inequality and polarization measurement for ordinal health variables. Section "Conclusion" includes some final remarks.

Cardinal Variables and the Measurement of Inequality in Health

This section presents inequality indices for cardinal health variables. We distinguish between the univariate and the bivariate approach and between individual-level health variables and group-level health data.[3]

The Univariate Approach to Measuring Health Inequality

The simplest case: cardinal health variables are available at the individual level

[3] Note that Erreygers and Van Ourti (2011a) present a matrix indicating which inequality index may be used for each type of health variables, making a distinction between ordinal, cardinal, ratio-scale, and fixed measurement scales.

Let h be a vector whose elements h_i refer to some cardinal measure of the health of individual i. A higher value of h_i means better health. We denote $\bar{h}$ as the average health level. Assume there are n individuals in the population analyzed and call e' a row vector whose elements are all equal to $(1/n)$. Let now G be a squared n by n matrix, called G-matrix (see, Silber, 1989), whose typical element g_{ij} will be equal to 0 if $i = j$, to -1 if $j > i$, and to $+1$ if $i > j$. Finally, let us also define a column vector s whose typical element $s_{i.}$ is equal to $\left(h_i / \sum_{i=1}^{n} h_i\right)$, these elements being ranked by decreasing health level (that is, by decreasing values of h_i).

We can now measure inequality in health as we measure income inequality and define, for example, the Gini index of health inequality as

$$I_G = e'Gs \tag{7.1}$$

Call now $r_i = (i/n)$ the relative (or "fractional") rank of individual i.

One can then show that if $n \to \infty$,

$$I_g \to \left[1 - \left(\frac{2}{n\bar{h}}\right) \sum_{i=1}^{n} h_i r_i\right] \tag{7.2}$$

If, on the contrary, we measure inequality in ill-health, that is, if a higher h_i refers to worse health, the Gini index of ill-health will be expressed as (see, Wagstaff & Van Doorslaer, 2004)

$$I_g \to \left[\left(\frac{2}{n\bar{h}}\right) \sum_{i=1}^{n} h_i r_i\right] - 1 \tag{7.3}$$

This Gini index may be given a simple graphical interpretation. Plot on the horizontal axis the cumulative values of $(1/n)$ and on the vertical axis plot the cumulative values of the shares s_i defined previously, these shares being ranked by increasing values of h_i (where h_i refers to good health). We then obtain a curve that will start at point (0,0) and end at point (1,1). This curve is a Lorenz curve and, as is well-known, the Gini index of health inequality is equal to twice the area lying between this Lorenz curve and the diagonal (the 45 degrees line defined previously).

The case of cardinal health variables available at the group level

Often health variables are not available at the individual level but only for given subgroups of the population. This is, for instance, the case of the infant mortality rate that, by definition, cannot be measured at the individual level, but at the regional level or for given socioeconomic groups.

Let m_j refers to the infant mortality rate in region j and let q_j represents the share of infants in region j in the total population of infants in the country. The infant mortality rate m for the country as a whole will then be defined as

$$m = \sum_j q_j m_j \tag{7.4}$$

The Gini index measuring the inequality in regional infant mortality rates will then be expressed as

$$I_{G,\text{infant mortality}} = [\ldots q_j \ldots]' G [\ldots (q_j (m_j/m)) \ldots] \tag{7.5}$$

where $[\ldots q_j \ldots]'$ is a row vector giving the share of the various regions in the total infant population of the country while the vector $[\ldots (q_j (m_j/m)) \ldots]$ is a column vector giving the share of the various regions in the infant mortality of the country and G is the G-matrix previously defined. Note that the elements of the vectors $[\ldots q_j \ldots]'$ and $[\ldots (q_j (m_j/m)) \ldots]$ are both ranked by decreasing values of the infant mortality rates m_j.

The case of variables for which only the expected distribution is available

There are finally variables which are not available at the individual level but their expected distribution is known. A life table, for example, gives the probabilities that an individual who is x years old will die at age $y > x$.

The life table provides then the distribution of deaths by age, that is, it gives for each age x the number of individuals $d(x)$ who are likely to die at this age, out of, say, 100,000 newborns.

Life expectancy at birth e_0 is then defined as

$$e_0 = \sum_x [xd(x)]/100,000. \quad (7.6)$$

Similarly, the inequality of the distribution of the ages at death may be measured via the Gini index $I_{G,ages\,at\,death}$ where, using the same kind of notation as before,

$$I_{G,\text{ages at death}} = \left[\cdots\left(\frac{d(x)}{100,000}\right)\cdots\right]' G \left[\cdots\left(\frac{xd(x)}{\sum_x xd(x)}\right)\cdots\right] \quad (7.7)$$

the elements in both the row and column vectors being ranked by decreasing age.

Decomposition

Assuming that health depends on a number of factors, a regression-based decomposition of the Gini coefficient will indicate the factors contributing to overall health inequality. Similarly, in a bivariate perspective, the concentration index can easily be decomposed into factors. Because the use of the concentration index is more common than that of the Gini coefficient in the empirical literature and because the decompositions of these indices are rather similar, we do not provide details about the decomposition of the Gini index here, but explain the decomposition of the concentration index in details below.

Empirical studies

Some studies employ the Gini coefficient to assess the level of health inequalities. Le Grand (1987) is one of the first to use (univariate) inequality indices to investigate health inequalities. Specifically, he focuses on age-at-death in a number of developed countries at the turn of the 1980s, and computes the Gini score, the absolute mean difference, and the Atkinson index in each country. He ranks countries by health inequality levels and finds that rankings are fairly stable between the three indices.

More recent works include a study on the role of lifestyle in inequality in premature mortality in Great Britain (Balia & Jones, 2008). Data come from the British Health and Lifestyle Survey (1984–1985) and its longitudinal follow-up (2003) that was traced to NHS Central Register to get information on deaths. The authors first compute predicted mortality from a probit model and transform it to get a positive variable. They then compute the Gini coefficient of predicted mortality. The decomposition analysis shows that lifestyles (smoking and sleeping pattern) strongly contribute to these inequalities.

For low-income countries, Tranvåg et al. (2013) employ the Gini (together with the concentration index and other indices) to measure inequality in length of life in Ethiopia. Data come from the 2000 and 2011 Ethiopian Demographic and Health Survey and the 2010 Global Burden of Disease study. While life expectancy increased over time, total length of life inequality (measured by the Gini coefficient) decreased during the same period. There is some heterogeneity between population groups, since inequalities (measured by the Gini score) are greater among males, rural and poorer individuals, than among females, urban and wealthier persons. Moreover, Van Malderen et al. (2019) analyze under-five mortality using data from the Demographic and Health Surveys (DHS) in 32 sub-Saharan countries in 2010–2016. The authors use a regression model to predict death rates and compute the Gini coefficient of these rates. While the main determinants of inequality in mortality vary between countries, a decomposition analysis reveals that the mother's education, child gender, household wealth, and the place of residence matter in a number of countries. Other very recent studies using the Gini coefficient to capture overall health inequalities include the article of Ikilezi et al. (2020) on vaccination for DTP3 in 45 sub-Saharan countries.

The Bivariate Approach to Health Inequality Measurement

The univariate approach does not capture the social dimension of inequalities. In this section, we thus focus on the bivariate approach that links social position and health.

Let y be a vector whose elements y_i are some measure of the standard living of individual i (e.g., his/her income). We define the concentration index C_G as

$$C_G = e' G \tau \tag{7.8}$$

where τ is a column vector of the shares s_i defined previously, these shares being now classified by decreasing values of the incomes y_i (rather than by decreasing values of the health levels h_i, as was the case in (7.1)).[4]

It can be shown (O'Donnell et al., 2007) that C_G is also expressed as

$$C_G = \left(\frac{2}{\bar{h}}\right) Cov(h_i, \rho_i) \tag{7.9}$$

where ρ_i is the fractional rank (i/n) of individual i in the distribution of the standards of living y_i, the individuals being ranked this time by increasing values of y_i.

Graphical interpretation

Here again it is possible to give a graphical interpretation, called the concentration curve, to this concentration index.

Plot on the horizontal axis the cumulative values of $(1/n)$. On the vertical axis plot now the cumulative values of the shares s_i, the latter being now classified by increasing values of the incomes y_i (rather than by increasing values of the health levels h_i). One then obtains again a curve that will start at point (0,0) and end at point (1,1). It can be shown that if this curve lies mostly under the 45 degrees line joining the points (0,0) and (1,1), the concentration index C_G will be positive, indicating that, as a whole, health increases with the standard of living. If, on the contrary, this curve lies mostly above the 45 degrees line, C_G will be negative, indicating that health decreases with the standard of living.

C_G will be equal to 0 either when all individuals have the same health level, whatever their standard of living, or when the sum of the areas lying below the 45 degrees line is exactly equal to the sum of the areas lying above the 45 degrees line (the concentration curve, although increasing, can clearly cut several times the 45 degrees line). It can be proved that this concentration index C_G is in fact equal to the sum of the areas lying between the concentration curve and the 45 degrees line, the areas below the 45 degrees line being given a positive sign and those above this line being given a negative sign.

[4] Note that the Gini coefficient and the concentration index give information on health attainment (Apouey & Silber, 2016).

Here also, when n is big enough ($n \to \infty$) the concentration index C_G may be expressed (see, Wagstaff & Van Doorslaer, 2004) as

$$C_G = \left[\left(\frac{2}{n\bar{h}} \right) \sum_{i=1}^{n} h_i \rho_i \right] - 1 \tag{7.10}$$

Decomposition

Importantly, a decomposition analysis may be applied to the concentration index to examine factors contributing to social health inequality (Wagstaff et al., 2003). Specifically, in this econometric approach, the concentration index can be decomposed into the sum of the contributions of the determinants of individual health. We first assume that individual health depends on a number of regressors x_k:

$$h_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i$$

where β_k captures the coefficient of regressor x_k (i.e., the "impact" of this regressor on health) and ε_i is the error term. Consequently, the concentration index can be rewritten as:

$$C_G = \sum_k \left(\frac{\beta_k \bar{x}_k}{\bar{h}} C_k \right) + \frac{GC_\varepsilon}{\bar{h}}$$

where $\bar{x}_k$ denotes the mean of x_k, C_k is the concentration index of x_k (which is defined in the same way as the concentration index for health), $\bar{h}$ is the mean of h. Finally, GC_ε is the generalized concentration index related to the residual ε_i. This is defined as $GC_\varepsilon = \frac{2}{n} \sum_{i=1}^{n} \varepsilon_i R_i$ where R_i is the fractional rank of individual i in the income distribution.

The equation $C_G = \sum_k \left(\frac{\beta_k \bar{x}_k}{\bar{h}} C_k\right) + \frac{GC_\varepsilon}{\bar{h}}$ is made up of two components: an explained component $\left(\sum_k \left(\frac{\beta_k \bar{x}_k}{\bar{h}} C_k\right)\right)$ and an unexplained component $\left(\frac{GC_\varepsilon}{\bar{h}}\right)$. The explained component is a weighted sum of the concentration indices of the regressors (C_k), in which the weights are $\frac{\beta_k \bar{x}_k}{\bar{h}}$ (note that β_k is the effect of the x_k factor on health, i.e., the size of the health change associated with a one-unit change in the x_k factor). The concentration index of each regressor (C_k) captures the level of inequality in the k factor across the socioeconomic distribution. The unexplained component is the part of the concentration index C_G that cannot be explained by variations in the x_k regressors across the socioeconomic distribution.

This decomposition thus assesses the role of different factors in the level of social health inequality, at one point in time, in a specific country, for instance. However, Wagstaff et al. (2003) also develop a decomposition of the *change* in the concentration index over time and of the *difference* in the concentration index between countries. To do this, they apply an Oaxaca-type decomposition (Oaxaca, 1973).[5]

Empirical illustrations using the concentration index

We here report some empirical illustrations using the concentration index. Note that some papers compute both the Gini and the concentration indices, to capture univariate and bivariate inequalities. Jürges (2010) focus on 11 European countries (not including Portugal) and the United States and computes a (sex-age standardized) Gini score and concentration index, for a continuous physical health variable. Interestingly, he tests whether his results depend on the choice of the stratification (i.e., socioeconomic) variable, and he alternatively uses education, income, and wealth. Evidence is mixed. Indeed, health inequalities are relatively high in the United States, England, or France, and low in Austria or Switzerland, independently of the stratification variable, but results for Italy significantly depend on the choice of the stratification variable. Simões et al. (2016) also focus on overall and social health inequality (as well as on health poverty and richness) but for Portugal in 2005–2006. After creating a cardinal health variable thanks to an algorithm, authors employ a number of indices, including the Gini coefficient and the concentration index. Comparing estimates with those presented by Jürges (2010) for other European countries and the United States, the article concludes that the Gini and the concentration indices for health are remarkably high in Portugal.

[5] Some features of social dispersion in health are not taken into account by inequality measures but may be better captured by polarization approaches. For this reason, building on the literature on bivariate health inequality (the concentration index) and univariate polarization (for cardinal variables), Apouey (2010) develops measures of bivariate polarization in health (for cardinal data). Like the concentration index, these measures can be decomposed into factors using a regression approach.

Some studies provide evidence for lower-income countries. For instance, to illustrate their decomposition methods, Wagstaff et al. (2003) study child malnutrition in Vietnam in 1993 and 1998. They compute the concentration index for the two years and show the decomposition of the index as well as the decomposition of the rise in inequality in malnutrition over time. Ataguba et al. (2016) compute the concentration index for a dichotomized self-assessed health variable, for South Africa. Good health is concentrated among the rich. The decomposition analysis reveals that social protection and employment, knowledge and education, and housing and infrastructure significantly contribute to inequality. Recent research also includes papers by Bado and Sathiya Susuman (2016) who compute the concentration index (as well as absolute and relative ratios) of under-five mortality rates according to the mother's education level, in selected sub-Saharan countries; Adesanya et al. (2017) who decompose the Gini and the concentration index for acute respiratory infection symptoms among young children in Nigeria between 2003 and 2013; Adeyanju et al. (2017) who explore socioeconomic inequalities in access to maternal and child healthcare in Nigeria in 1990 and 2008; Nkonki et al. (2011) who focus on the drivers of social inequality in child morality, HIV transmission, and vaccination, using a decomposition of the concentration index, for three sites in South Africa; and Mané (2013) who studies social inequalities in health care use in Senegal.

Corrections of the concentration index

While the concentration possesses a number of interesting features (visual representation, decomposition), some limitations have been highlighted (Erreygers, 2009): comparing population with different mean health levels is problematic because rankings depend on whether one focuses on ill-health or on good health, and the concentration index is somewhat arbitrary for qualitative health data. Erreygers (2009) derives a family of inequality indicators (the so-called corrected concentration index) to overcome these three limitations. This new measure should only be used for cardinal health variables that have finite lower and upper bounds. It satisfies four properties (transfer, level independence, cardinal invariance, and mirror) and can be easily decomposed.

Quentin et al. (2014) use the index proposed by Erreygers (2009) as well as the concentration index and other indices to assess inequality in child mortality in ten African countries, exploiting data from the Demographic and Health Surveys. While these inequalities exist in all cities, there are important differences in inequality levels and evolutions between cities.

While binary health variables are common (e.g., having chronic condition or not), they complicate the measurement of inequality. A series of papers discuss the corrections to the concentration index for this type of health variables (Erreygers & Van Ourti, 2011b; Kjellsson & Gerdtham, 2013; Wagstaff, 2005, 2011a, 2011b).

The empirical literature employs these corrected indices for dummy variables. For instance, Van Malderen et al. (2013) investigate the determinants of overall and social inequality in under-5 mortality in 13 African countries. Data come from the Demographic and Health Surveys between 2007 and 2010. The authors compute the Gini index (to measure overall health inequality) and the normalized Erreygers concentration index for binary outcomes (to measure social health inequality). The decomposition analysis indicates that the birth order and interval as well as the region matter for overall inequality. Social health inequality is significant in five countries (Democratic Republic of Congo, Egypt, Madagascar, Nigeria, and Sao Tome and Principe) and the main determinants of this inequality are household wealth, father's occupation, and mother's education. Moreover, Ataguba et al. (2011) apply a concentration index, generally normalized using the work of Wagstaff (2005), to compute social inequalities in dichotomous variables (self-reported illness and disability), for the South African General Household Survey, between 2002 and 2008.

The Univariate Approach to Measuring Health Achievements

The case of variables for which only the expected distribution is available

Following work by Kolm (1969) and Atkinson (1970), Sen (1973) suggests an index of welfare combining per capita income and income inequality. This index turns out to correspond to the concept of "equally distributed equivalent level of income" proposed by Atkinson (1970), a notion identical to that of "equal equivalent income" defined by Kolm (1969). The "equally distributed equivalent level of income" y_E is in fact equal to the product of the average income $\overline{y}$ and the complement to one of the inequality index (e.g., the Gini or the Atkinson indices of inequality) which can be derived from the social welfare function selected.[6]

The same kind of idea may be applied to health (e.g., life tables). Life expectancy (e_0) is in fact the mean of the distribution of years lived. It does not take into account the dispersion of such a distribution. If we want to include this aspect, we may then define a new indicator that will "penalize" the mean, the more dispersion there is. This is what Silber (1983) does when he applies this idea to life tables. Denoting an inequality index in health by I, Silber defines what he calls the "equivalent length of life (E. L. L)" where

$$E.L.L. = e_0(1 - I) \tag{7.11}$$

Note that e_0 was originally introduced as a measure of development. Indeed, while Hicks and Streeten (1979) recommend using life expectancy as a measure of development (because it "… would be a good single measure of basic needs"), Silber (1983) extends these ideas by proposing to use the concept of E. L. L. as a measure of development.

But this concept can naturally be used also as a measure of health achievement. This idea may in fact be applied to any indicator of health. In other words, a measure of health achievement should be an increasing function of the average level of the health indicator selected and a decreasing function of the degree of inequality of the distribution of this health indicator.

This implies that in computing such a measure of health achievement, the weight of an individual will be higher, the lower the value for this individual of the health indicator selected.

A graphical interpretation

Plot on the horizontal axis the cumulative population shares. On the vertical axis multiply the cumulative income shares used in defining a Lorenz curve by the average income. We will then obtain what has been called a *Generalized Lorenz curve* (see, Shorrocks, 1983). Such a curve

[6] Such an inequality index should vary between 0 and 1, which, for example, is not the case of the so-called Theil (1967) indices.

starts at point (0,0) and ends at point $(1, \overline{y})$ where, as before, $\overline{y}$ is the average income.

Since the area lying between the diagonal and a Lorenz curve is equal to half the Gini index I_G, the area lying between a Generalized Lorenz curve and a line starting at point (0,1) and ending at point $(1, \overline{y})$ will be equal to half the product $\overline{y} I_G$. so that the area lying below the Generalized Lorenz curve will be equal to half the product $\overline{y}(1 - I_G)$ which is in fact identical to half the equally distributed equivalent level of income y_E.

One can naturally apply the concept of Generalized Lorenz curve to measure the welfare derived from some health achievement. Such a welfare measure would in fact give a greater weight to an individual, the lower the level of his/her health.

One may however think of an alternative approach, one where the weight of an individual, when measuring the welfare derived from his/her health, would be higher, not the lower the level of his/her health, but the lower his/her income. This is in fact the approach taken by Wagstaff (2002) in his definition of health achievement.

We can call such an approach the pro-poor approach to the measurement of health achievements.

A Pro-poor Approach to the Measurement of Health Achievement

Using what was defined previously as the bivariate approach to health inequality measurement, Wagstaff (2002) proposes to define health achievement as the weighted average of the health levels of the various individuals, the weights being higher, the poorer the individual.

Let $\overline{h}$ be the average value of the health indicator and let again C_G be the concentration ratio of the health indicator (with respect to income). The pro-poor measure of health achievement $A_{h,PP}$ is then expressed as

$$A_{h,PP} = \overline{h}(1 - C_G) \tag{7.12}$$

A graphical interpretation

In the same way as we derived previously a Generalized Lorenz curve, we can now define the concept of Generalized Concentration curve. We simply have to order the vertical coordinates of the Generalized Lorenz curve not by increasing values of the health variable, but by increasing income. It is then easy to derive that the area under such a Generalized concentration curve will be equal to half the product $\overline{h}(1 - C_G)$, that is, to half the pro-poor health achievement indicator $A_{h,PP}$.

The Measurement of Inequality in Health Opportunities

The Univariate Approach to Defining Indices of Inequality of Opportunities

Assume we have data on the probability of access to a given service (e.g., drinkable water) by population subgroups (e.g., regions or ethnic groups). Let $M_{i.}$ represent the total number of individuals in subgroup i and let the subscript j indicate whether the group does not have access ($j = 1$) or has access ($j = 2$) to the service. Let now M_{i1} and M_{i2} represent, respectively, the number of individuals in subgroup i who do not have access and have access to the service, with $M_{i1} + M_{i2} = M_{i.}$. Let also M refer to the total population, with $M = \sum_i M_{i.}$ and $M_2(M_1)$ the total number of individuals in the population having access (and not having access) to the service, with $M_2 = \sum_i M_{i2}$ (and $M_1 = \sum_i M_{i1}$).

If there was independence between the probability of belonging to a given ethnic group (region) i and that of having access to a specific service j, the product $(M_2/M)(M_{i.}/M)$ would be equal to (M_{i2}/M).

If this is not the case for every population subgroup, there is no equality of opportunities between the subgroups and such an inequality may be measured via the use of traditional income inequality indices. Various inequality indices may be used to measure such a gap between "predicted" and "actual" shares, such as the Gini or the Theil indices.

Call m_{ij} the ratio $\left(M_{ij}/M\right)$ and $m_{i.}$ and $m_{.j}$ the ratios $(M_{i.}/M)$ and $\left(M_{.j}/M\right)$. Therefore, if there are only two possibilities (having or not having access to a specific health service), the Theil indices of inequality of opportunities (of access to this health service) will then be expressed as

$$T_1 = \sum_i (m_{i.} m_{.2}) ln \left[\frac{(m_{i.} m_{.2})}{(m_{i2})} \right] \tag{7.13}$$

or as

$$T_2 = \sum_i (m_{i2}) ln \left[\frac{(m_{i2})}{(m_{i.} m_{.2})} \right] \tag{7.14}$$

Similarly, using previously defined notations, the Gini index of inequality of opportunities (of access to the health service) will be expressed as

$$I_G = [\ldots(m_{i.}m_{.2})\ldots]' G[\ldots(m_{i2})\ldots] \tag{7.15}$$

In short, $(m_{i.}m_{.2})$ represents the "a priori" probability for group i of having access to the service while (m_{i2}) represents the "a posteriori" probability for group i of having access to this service. In (7.15), the elements of the vectors $[\ldots(m_{i.}m_{.2})\ldots]'$ and $[\ldots(m_{i2})\ldots]$ are both ranked by decreasing values of the ratios $(m_{i2}/(m_{i.}m_{.2}))$.

Naturally, we could use another inequality index, such as the Dissimilarity index D (relative mean deviation) where

$$D = \sum_i (m_{i.}m_{.2}) \left| \frac{(m_{i2})}{(m_{i.}m_{.2})} - 1 \right| \tag{7.16}$$

D may be also expressed as

$$D = \sum_i |(m_{i2}) - (m_{i.}m_{.2})| \tag{7.17}$$

Call now p_i the ratio $\left(\frac{M_{i2}}{M_{i.}}\right)$ and p the ratio $\left(\frac{M_{.2}}{M}\right)$. We may then write D as

$$\begin{aligned} D &= \sum_i \left| \left(\frac{M_{i2}}{M_{i.}}\right)\left(\frac{M_{i.}}{M}\right) - \left(\frac{M_{i.}}{M}\right)\left(\frac{M_{.2}}{M}\right) \right| \\ &= \sum_i \left(\frac{M_{i.}}{M}\right) \left| \left(\frac{M_{i2}}{M_i}\right) - \left(\frac{M_{.2}}{M}\right) \right| \\ &= \sum_i m_{i.}|p_i - p| \end{aligned} \tag{7.18}$$

In other words, the relative mean deviation is a weighted average of the absolute gaps between the share of individuals having access to the service in group i and the corresponding share in the whole population, the weights of the groups being the population share in the total population. This measure is in fact the one adopted by Paes de Barros, Ferreira, et al. (2008) and Paes de Barroa, Molinas Vega et al. (2008).

The Bivariate Approach to Measuring Inequality of Opportunities

We have hitherto shown how to measure, for example, inequality in the access to a physician for a delivery, no matter what the socioeconomic background of the individual is. Let us however assume that we want to analyze the link which exists between this access to a physician and the socioeconomic background of the individual.

In such a case, we would, in the equation defining the Gini index on the basis of the G-matrix, classify the "a priori" probabilities $(m_{i.}m_{.2})$ and the "a posteriori" probabilities (m_{i2}) in (7.16) not by decreasing ratios $\left[\frac{(m_{i2})}{(m_{i.}m_{.2})}\right]$ but by decreasing socioeconomic background. We would then compute not the Gini index but the concentration ratio of access to the health service.

The Concept of Human Opportunity Index

Defining "Welfare-Related" Indices of Opportunity

In recent work, Paes de Barros, Ferreira, et al. (2008) and Paes de Barroa, Molinas Vega et al. (2008) define what they call the Human Opportunity Index *HOI* which, using previous notations, is defined as

$$\mathrm{HOI}_D = 0.5p(1 - D) \tag{7.19}$$

It is easy to observe that HOI increases with the prevalence of opportunities (an increase in p increases HOI) as well as with an improvement in the way opportunities are allocated (a reduction in D increases HOI).

It is naturally possible to derive a similar expression using the Gini index and define a human opportunity index HOI_G as

$$\mathrm{HOI}_G = p(1 - G) \tag{7.20}$$

The idea here is simply to draw a Generalized Lorenz curve of the probabilities of having access to the health service.

Defining "Pro-poor" Human Opportunity Indices

One can think of another way of defining a "Human Opportunity Index," one that would give more weight to individuals coming from a lower socioeconomic background. In order to derive such indices, we will now have to rank the "a priori" and "a posteriori" probabilities according to the socioeconomic background of the individuals. In other words, we want here to derive "Human Opportunity Indices" on the basis of the Concentration index rather than on the basis of Gini inequality index.

The "Gini-related Pro-Poor Human Opportunity Index" (HOI_{PP}) will then be defined as

$$\text{HOI}_{PP} = p(1 - C_G) \tag{7.21}$$

A graphical interpretation

Let us rank the cumulative values of the "a priori" probabilities that are plotted on the horizontal axis and the cumulative values of the "a posteriori" probabilities that are plotted on the vertical axes to derive a kind of Generalized Lorenz curve, where these two sets of cumulative probabilities are not ranked by increasing values of the ratios of the "a posteriori" over the "a priori" probabilities for the individuals of having access to the service, but by increasing values of the socioeconomic background of these individuals. We will then obtain a curve which we can call the "Generalized Gini Concentration curve" of the probabilities for the individuals of having access to the health service.

It is then easy to prove that the area lying under such a curve is in fact equal to half the product $0.5p(1 - C_G)$.

Measuring Health Inequality and Polarization with Ordinal Variables

Partial Orderings

In a path breaking paper, Allison and Foster (2004) start by stressing the fact that traditional measures of inequality, such as the Gini coefficient, the index of Atkinson (1970), the entropy related measures introduced by Theil (1967) or the variance, are "mean based" because either inequality is viewed as deviation from the mean or these indices are normalized via the mean. However, when working with ordinal variables such as self-assessed health, this reference to the mean becomes problematic as the notion of mean in such a case is not well defined. They give an empirical illustration showing that, depending on the scale selected for the different ordinal categories, one can draw different conclusions when comparing health inequality in different populations.

There are however cases where such comparisons cannot lead to a reversal of ranks, whatever the scale used. Such an unambiguous ranking occurs when one distribution first-order dominates the other one. Allison and Foster then wonder what reference point should be chosen if the mean cannot be used with ordinal variables. They recommend using the median and define then the concept of "spread away from the median." Assume two distributions $\{d_i\}$ and $\{d_i^{'}\}$ of health levels. We can then state that $\{d_i^{'}\}$ has a greater spread than $\{d_i\}$ (which Allison and Foster express as $\{d_i\}S\{d_i^{'}\}$) if

- $\{d_i\}$ and $\{d_i^{'}\}$ have the same median category m
- For all health levels $i < m$, we have $D_i \geq D_i^{'}$
- For all categories $i \geq m$, we have $D_i^{'} \geq D_i$

In short, $\{d_i^{'}\}$ first order dominates $\{d_i\}$ below the median, while $\{d_i\}$ first order dominates $\{d_i^{'}\}$ for the median category and above. This is the essence of what Allison and Foster call "median preserving spread."

Let us take a simple numerical illustration. Let the absolute frequencies of $\{d_i\}$ be: $\{2 \ 3 \ 8 \ 4 \ 3\}$ and those of $\{d_i^{'}\}$: $\{1 \ 2 \ 14 \ 2 \ 1\}$. The relative frequencies for $\{d_i\}$ are then $\{0.1 \ 0.15 \ 0.4 \ 0.2 \ 0.15\}$ and those for $\{d_i^{'}\}$: $\{0.05 \ 0.10 \ 0.70 \ 0.10 \ 0.05\}$. The corresponding cumulative frequencies $F\{d_i\}$ and $F\{d_i^{'}\}$ are then respectively: $F\{d_i\} = \{0.10 \ 0.25 \ 0.65 \ 0.85 \ 1\}$ and $F\{d_i^{'}\} = \{0.05 \ 0.15 \ 0.85 \ 0.95 \ 1\}$ so that clearly $F\{d_i^{'}\}$ lies below $F\{d_i\}$ before the median and above it at the median and beyond it.

What Allison and Foster (2004) then suggest is to draw two new curves which beyond the median will be the same as the two curves derived from the cumulative frequencies $F\{d_i\}$ and $F\{d_i^{'}\}$, while for the health levels lower than the median, these cumulative frequency curves will be "flipped" (for more details and a graphical illustration, see Allison & Foster, 2004).

Allison and Foster then draw our attention to the fact that the ranking $\{d_i\}S\{d_i^{'}\}$ that was defined previously is only a partial ordering of distributions. This ordering is indeed reflexive (eSe for any distribution e), transitive (dSe and eSf imply dSf) but it is not complete" (because very often neither dSe nor eSd holds.

Allison and Foster then proceed to define what they call the "*S*-curve" which is a graphical representation of the partial ordering *S*. In the first stage of the derivation of this curve, they ask us to "flip over to the left" the portion of the cumulative distributions that lie to the right of the median level of health. In a second stage, the curves that appear to the left of the median are then rotated 90 degrees to obtain what the authors call the *S*-curve. They then draw our attention to the fact that the base of this *S*-curve corresponds in fact to the range of the population having the median level of health. On both side of this range, we have the groups that are respectively one and two health levels away from the median. It should then be clear that if, say, the *S*-curve for $\{d_i\}$ lies inside the *S*-curve for $\{d_i^{'}\}$, this would indicate that $\{d_i\}$ is more "spread away" from the median than $\{d_i^{'}\}$ as far as the distribution of happiness levels is concerned.

Note that, as stressed by Kobus et al. (2019), the approach of Allison and Foster is a special case of a relationship developed by Mendelson (1987) in his paper on quantile preserving spreads.

Inequality and Polarization Indices

Noting that the median plays a central role in the analysis of Allison and Foster (2004), Apouey (2007) reminds the reader that the literature on polarization (see, Duclos et al., 2004; Esteban & Ray, 1994; Wolfson, 1994) emphasizes also the median. This literature focuses on cardinal variables and emphasizes two bi-polarization principles: that of "increasing spread" and that of "increased bipolarity."

The first principle (increasing spread) states that moving from the middle position (the median) to the tails of the distribution will make the distribution more polarized. Taking the income distribution as illustration, this means that a rank preserving increment in incomes above the median or a rank preserving reduction in income below the median will widen the distribution, that is, extend the distance between the two groups (those above and below the median) and hence increase the degree of bi-polarization (the rich become richer and the poor poorer).

The second principle (increased bipolarity) concerns on the contrary the case where the incomes below the median or those above the median become closer to each other. This implies that there has been some "bunching" of the two groups so that the gaps between the incomes below the median (or those above the median) have been reduced. In such a case, bi-polarization is assumed to increase.

There is hence a fundamental difference between the concepts of “inequality” and “bi-polarization”: any regressive transfer will increase inequality but it will increase the degree of bi-polarization only if this transfer takes place across the median. On the contrary, it will decrease bi-polarization if it takes place on the same side of the median.

The approach of Apouey (2007) uses these two principles in deriving polarization indices for the case of ordinal variables. In addition, she interprets the principle of transfer by assuming that, in the case of ordinal variables, it refers to a movement of individuals from one category to another.

Taking self-assessed health, an ordinal variable, as illustration, Apouey assumes then that there is no polarization when everyone is in the same health category, while polarization reaches a maximum when half of the population is in the lowest category and half in the highest.

She then derives axiomatically the following index I_{AP}:

$$I_{AP} = 1 - \frac{2^{\alpha}}{I-1} \sum_{i=1}^{I-1} |P_i - 0.5|^{\alpha} \qquad (7.22)$$

where P_i refers to the cumulative frequency for health state i. Apouey suggests to calibrate α in such a way that the index I_{AP} will be equal to 0.5 for a uniform distribution (same number of individuals in each health category). In other words, such a uniform distribution is assumed to be an intermediate state between the cases of minimum and maximum polarization that were defined previously.

Apouey indicates that the index I_{AP} has the following properties:

- *Continuity*: Small changes in the distribution do not lead to a large variation in the index
- *Slide*: This assumption implies that the polarization index takes only into account the order of the proportions in the categories, and not the corresponding welfare levels (see Apouey [2007] for an exact definition).
- *Symmetry*: The polarization index is symmetric since it does not change when the categories are ranked in the reverse order.

- *Distance*: When the population is divided into two clusters, if the distance between the clusters is shortened, the polarization index decreases.
- *Frequency*: When the population is divided into two groups, then the polarization index will decrease when the distribution of individuals between the two peaks diverges from half and half.

Abul-Naga and Yalcin (2008) also develop indicators of dispersion for ordinal variables. They present their indicators as "inequality" indicators. However, as in the polarization literature, their indicators indicate that "inequality" is largest when half of the population is in the lowest category and the other half is in highest one, like in Apouey (2007). For this reason, we believe that the distinction between "inequality" and "polarization" in the literature on ordinal outcomes is unclear and requires additional thought.

The main goal of Abul-Naga and Yalcin (2008) is to characterize the entire class of inequality indices founded on the ordering defined by Allison and Foster and satisfying the following properties: continuity, scale invariance, normalization, and aversion to median preserving spreads. Continuity requires that small changes in the distribution of the variables or in the health scale do not produce large jumps in the value taken by the inequality index. Scale invariance assumes that if two distributions have the same level of inequality under a given health scale, they should be considered as equivalent, whatever the selected scale. Normalization implies that, whatever the health scale, inequality will be minimal if every individual is located at the median health state. Abul Naga and Yalcin then assume that this minimum level of inequality is zero. Finally, the axiom of aversion to median preserving spreads says that if a distribution *d'* is derived from a distribution *d* by a median preserving spread, *d'* should be less equal than *d*.

Let p_i be the proportion of individuals having health *i*. Assume that the various categories are ordered by increasing health status and define P_i as the cumulative values of the probabilities p_i, $(P_i = p_1 + p_2 + \ldots + p_i)$. Moreover, let *m* denote the median health state. Abul Naga and Yalcin first define the ordinal inequality index I_{AY} with

$$I_{AY} = 1 - \left[\frac{\left(2 \sum_{i=1}^{I} |P_i - 0.5| - 1 \right)}{(I - 1)} \right] \tag{7.23}$$

The index I_{AY} obeys the four axioms postulated by these authors. They then characterize a more general family of indices that satisfy these four axioms and define the following index:

$$I_{AY}^{\alpha,\beta} = \frac{\sum_{i<m} (P_i)^{\alpha} \sum_{i \geq m} (P_i)^{\beta} + (I + 1 - m)}{k_{\alpha,\beta} + (I + 1 - m)} \tag{7.24}$$

with

$$k_{\alpha,\beta} = (m - 1)\left(\frac{1}{2}\right)^{\alpha} - \left[1 + (I - m)\left(\frac{1}{2}\right)^{\beta}\right] \tag{7.25}$$

The authors prove that the index defined in (7.24) satisfies the axioms of continuity, aversion to median preserving spreads, a slightly stronger version of the normalization axiom (see, Abul Naga & Yalcin, 2008), and an axiom of scale independence. This last axiom implies that the inequality index will not vary, as long as one selects increasing health scales.

Note that when $\alpha = \beta = 1$ in (7.24) and (7.25), one obtains the index I_{AY} defined in (7.23). This index I_{AY} is symmetric which means that equal deviations from 0.5 below and above the median lead to the same value of the inequality index.

The generalized formulation given in (7.24) and (7.25) allows one to introduce asymmetry, that is, to get different results when deviations from 0.5 take place below or above the median. For example, assume that $\alpha \geq 1$ and $\beta \geq 1$. Then for a given value of β, the index $I_{AY}^{\alpha,\beta}$ becomes more sensitive to the cumulative probability mass at the bottom of the distribution as $\alpha \to 1$. On the contrary, as $\alpha \to \infty$, the index index $I_{AY}^{\alpha,\beta}$ will ignore the dispersion below the median. Similar considerations hold evidently when varying β for a given value of α.

Kobus and Milos (2012) derive a class of inequality measures (for ordinal data) which are decomposable by population subgroups.

This leads to a severe restriction on the functional forms of inequality/polarization indices. More precisely, Kobus and Milos (2012) start by following Abul Naga and Yalcin (2008) and assume that an inequality index for ordinal variables satisfies the following axioms:

- Continuity
- Scale invariance: the ordering of distributions derived from an index does not vary when the scale changes.
- Scale independence: the index does not depend on the scale. Note that scale independence implies scale invariance.
- Normalization: for the most equal distribution the index has a zero value while a value of 1 is assigned to the most unequal distribution.
- The "EQUAL" assumption: the index is consistent with Allison and Foster equality ordering.

In addition, the authors require the index to be decomposable, that is, it can be represented as a function of the inequality in the different subgroups and of the subgroup relative sizes.

Kobus and Milos (2012) then prove that an index fulfills the axioms of continuity, normalization, scale independence and decomposability if and only if it is of the form

$$I = G\left(\sum_{i=1}^{I} p_i\right) \tag{7.26}$$

If one prefers to impose only scale invariance rather than the stronger assumption of scale independence, the authors show that an index fulfills the axioms of continuity, normalization, scale invariance and decomposability if and only if it is of the form

$$I = G\left(\sum_{i=1}^{I} a_i\, p_i, c\right) \tag{7.27}$$

where c refers to the scale and the coefficient a_i is the weight assigned to the probability that an individual belongs to category i.

If, in addition to the assumption of continuity, normalization, scale independence, and decomposability we impose the "EQUAL assumption", then the index will have the form given in (7.26) with

- G a strictly increasing function and $a_i \geq a_{i+1}$ when $i < m$ and $a_i \leq a_{i+1}$ when $i \geq m$
- G a strictly decreasing function and $a_i \leq a_{i+1}$ when $i < m$ and $a_i \geq a_{i+1}$ when $i \geq m$.

Since the "EQUAL assumption" means that a distribution more concentrated around the median is more equal, the inequality index gives higher weight to a category, the further away it is from the median.

Kobus and Milos (2012) then propose the following generalization of the index introduced by Abul Naga and Yalcin (2008):

$$I_{a,b} = \frac{a \sum_{i<m} P_i - b \sum_{i \geq m} P_i + b(n+1-m)}{(a(m-1)+b(n-m))/2} \quad \text{with } a \geq 0; b \geq 0. \tag{7.28}$$

Note that when $a = b = 1$, Eq. (7.28) becomes identical to the index proposed by Abul Naga and Yalcin. It turns out that this is also true whenever a and b are identical.

Kobus and Milos (2012) however show that the only indices that are decomposable are the linear forms of the absolute value index I_{AP} introduced by Apouey (2007) which is identical to one of the indices proposed by Abul Naga and Yalcin (2008).

Lazar and Silber (2013) take a somehow different route and start by observing that the indices introduced by Reardon (2009) to measure ordinal segregation may also be used in other domains where only ordinal information is available.

Let P, as before, refer to the distribution function. Define now on the interval [0,1] a continuous function $f(P)$ with the following properties: it is increasing when $P \epsilon (0, 0.5)$ and decreasing when $P \epsilon (0.5, 1)$. Moreover, its value will reach a maximum at $P = 0.5$ so that $f(0.5) = 1$ and a minimum when $P = 0$ or $P = 1$ so that $f(0) = f(1) = 0$. Let, as before, I refer to the number of categories and define a function v as

$$v = \left(\frac{1}{I-1}\right) \sum_{i=1}^{I-1} f(P_i) \tag{7.29}$$

Reardon (2009) introduces then the following four functions $f(P)$:

$$f_1(P) = -[P \log_2 P + (1-P) \log_2 (1-P)] \tag{7.30}$$

$$f_2(P) = 4P(1 - P) \tag{7.31}$$

$$f_3(P) = 2\sqrt{P(1 - P)} \tag{7.32}$$

$$f_4(P) = 1 - |2P - 1| \tag{7.33}$$

Combining (7.29) with one of the four functions defined in Eqs. (7.30) to (7.33) yields indeed ordinal inequality indices that satisfy the four desirable properties stressed by Abul Naga and Yalcin (2008): continuity, scale invariance, normalization, and aversion to median preserving spreads.

Note also that the index proposed by Abul Naga and Yalcin (2008) in (7.23) corresponds in fact to the combination of (7.29) and (7.33). To check this, insert $f_4(P)$ in (7.25) and remember that $P_I = 1$. We then derive that

$$\begin{aligned}
v &= \left(\frac{1}{I-1}\right)\sum_{i=1}^{I-1}(1 - |2P_i - 1|) \\
&= \frac{(I-1) - \sum_{i=1}^{I-1}|2P_i - 1|}{I-1} \\
&= 1 - \frac{\sum_{i=1}^{I-1}|2P_i - 1|}{I-1} \\
\leftrightarrow v &= 1 - \frac{\sum_{i=1}^{I}|2P_i - 1| - (2P_i - 1)}{I-1} \\
&= 1 - \frac{\sum_{i=1}^{I}|2P_i - 1| - 1}{I-1} \\
&= 1 - \frac{2\sum_{i=1}^{I}|P_i - 0.5| - 1}{I-1}
\end{aligned} \tag{7.34}$$

Note that the last expression on the R.H.S. of (7.34) is in fact the index proposed by Apouey (2007) as well as Abul Naga and Yalcin (2008).

As mentioned previously, Abul Naga and Yalcin (2008) also derived a parametric family of inequality indices where different weights can be put on different parts of the distribution (see Eqs. (7.13) and (7.14)). Their measure may in fact be expressed as

$$I_{AY}^{\alpha,\beta} = \frac{f^{Maximum} - f^{Actual}}{f^{Maximum} - f^{Minimum}} \tag{7.35}$$

where $f^{Maximum}$, $f^{Minimum}$, and f^{Actual} refer, respectively, to the maximal, minimal, and actual value of some function f. More precisely

$$f^{Actual} = \sum_{i<m} (P_i)^{\alpha} - \sum_{i\geq m} (P_i)^{\beta} \tag{7.36}$$

while

$$f^{Minimum} = (I + 1 - m) \tag{7.37}$$

which corresponds to the most egalitarian distribution (every individual is located at the median)

and

$$f^{Maximum} = k_{\alpha,\beta} \tag{7.38}$$

where, as indicated in (7.25), $k_{\alpha,\beta} = (m-1)\left(\frac{1}{2}\right)^{\alpha} - \left[1 + (I-m)\left(\frac{1}{2}\right)^{\beta}\right]$ and corresponds to the least egalitarian distribution, where half of the population is in the worst health category and half in the best health category.

Lazar and Silber (2013) then extend this analysis by postulating two increasing and continuous functions, namely $g_1(P_i)$ for $i < m$, and $g_2(P_i)$ for $i \geq m$. More precisely, they postulate that $g_1\left(\frac{1}{2}\right) = 1 - g_2\left(\frac{1}{2}\right)$, a condition which guarantees continuity, and that $g_1(0) = 0$; $g_2(1) = 1$; $g_1(1/2) = 1$ so that $g_2(1/2) = 0$. These assumptions allow making a comparison with the approach of Reardon (2009) that was mentioned previously. The extension of the index I'_{AY} is then written as

$$I_{AY}^{\text{extended}} = \frac{\sum_{i<m} g_1(P_i) - \sum_{i\geq m} g_2(P_i) + (I-m+1)g_2(1) - (m-1)g_1(0)}{[(m-1)g_1(1/2) - (I-m)g_2(1/2) - g_2(1)] + [(I-m+1)g_2(1) - (m-1)g_1(0)]}$$

$$\leftrightarrow I_{AY}^{\text{extended}} = \frac{\sum\limits_{i<m} g_1(P_i) - \sum\limits_{i \geq m} g_2(P_i) + (I - m + 1)}{[(m-1)-1] + [(I-m+1)]}$$

$$= \frac{\sum\limits_{i<m} g_1(P_i) - \sum\limits_{i \geq m} g_2(P_i) + (I - m + 1)}{(I-1)} \tag{7.39}$$

Let us now define a general continuous function $f(P_i)$ as

$$f(P_i) = g_1(P_i) \text{ if } 0 \leq P_i < \frac{1}{2} \quad \text{and}$$
$$f(P_i) = 1 - g_2(P_i) \text{ if } \frac{1}{2} \leq P_i < 1. \tag{7.40}$$

We can then rewrite Reardon's (2009) function v as follows

$$v = \left(\frac{1}{I-1}\right) \sum_{i=1}^{I-1} f(P_i)$$

$$= \left(\frac{1}{I-1}\right) \left[\sum_{i=1}^{m-1} g_1(P_i) + \sum_{i=m}^{I-1} (1 - g_2(P_i))\right]$$

$$\leftrightarrow v = \left(\frac{1}{I-1}\right) \left[\sum_{i=1}^{m-1} g_1(P_i) - \sum_{i=m}^{I-1} g_2(P_i) + (I-m)\right]$$

$$\leftrightarrow v = \left(\frac{1}{I-1}\right) \left[\sum_{i=1}^{m-1} g_1(P_i) - \sum_{i=m}^{I} g_2(P_i) + g_2(1) + (I-m)\right]$$

$$\leftrightarrow v = \left(\frac{1}{I-1}\right) \left[\sum_{i=1}^{m-1} g_1(P_i) - \sum_{i=m}^{I} g_2(P_i) + (I-m+1)\right] \tag{7.41}$$

Let us now combine (7.32) and (7.39) to derive that

$$f_4(P) = 2P \text{ if } 0 \leq P < \frac{1}{2} \quad \text{and}$$
$$f_4(P) = 2(1-P) \text{ if } \frac{1}{2} \leq P \leq 1 \tag{7.42}$$

Add now the two parameters α and β introduced by Abul Naga and Yalcin (2008) and defined previously. This allows us to define a new function

$f_5(P)$ with

$$f_5(P) = (2P)^\alpha \text{ if } 0 \le P < \frac{1}{2} \quad \text{and}$$
$$f_5(P) = (2(1-P))^\beta \text{ if } \frac{1}{2} \le P \le 1 \tag{7.43}$$

with $\alpha > 0$ and $\beta > 0$.

Note that this function $f_5(P)$, like the four functions introduced by Reardon and defined previously in Eqs. (7.30) to (7.33), has the properties mentioned previously: it is increasing for $P \epsilon (0, 0.5)$, decreasing for $P \epsilon (0.5, 1)$, equal to 1 when $P = 0.5$ and to 0 when $P = 0$ or $P = 1$.

If we now combine (7.29) and (7.43) we obtain an index I_5 written as

$$I_5 = \left(\frac{1}{I-1}\right)\left\{\sum_{i=1}^{m-1}(2P)^\alpha + \sum_{i=m}^{I-1}[2(1-P)]^\beta\right\} \tag{7.44}$$

$$\leftrightarrow I_5 = \left(\frac{1}{I-1}\right)\left\{\sum_{i=1}^{m-1}(2P)^\alpha - \sum_{i=m}^{I}\left[1-(2(1-P))^\beta\right] + (I-m+1)\right\} \tag{7.45}$$

which is identical to the index I_{AY}^{extended} defined in (7.39).

Lv et al. (2015) propose a measure of the inequality of ordinal variables, taking self-assessed health as an illustration, derived in two stages. First, a measure of the inequality between any two different health outcomes is defined; then, these inequalities are aggregated via a simple weighted sum, in which the further apart the two health outcomes, the higher the weight attached to the inequality between these two health outcomes. Lv et al. (2015) derive axiomatically the two following indices:

$$I_{LWX1} = \sum_{i=1}^{I}\sum_{k \neq i}\left(\frac{2}{(I-1)}\right)|k-i| f_k f_i \tag{7.46}$$

where I is the number of possible health outcomes and f_k and f_i are the proportion of individuals with health outcomes k and i, respectively, and

$$I_{LWX2} = \sum_{i=1}^{I}\sum_{k \neq i}\alpha^{I-1-|k-i|} f_k f_i \tag{7.47}$$

with α = 0.9, 0.6, 0.3 or 0.1.

Lv et al. (2015) show that the two previous indices obey a certain number of axioms. In the case of health inequality, when self-assessed health is an ordinal variable, these axioms may be stated as follows:

- *Focus:* This property implies that any additional information about individuals, such as their gender or age, should not play any role in constructing an index of health inequality.
- *Additivity:* An index of health inequality should be the sum of all "individual" health inequalities. The measure sums up all possible inequalities of any two different health outcomes.
- *Independence:* This property requires that any change in the degree of health inequality between two health outcomes, h_m and h_j, as a consequence, say, of an increase in the frequency of health outcome, h_j, is independent of the health frequency of health outcome, h_j.
- *Perfect equality*: If everyone has the same health outcome, then health inequality is equal to zero.
- *Invariance to simple switches*: When all individual health outcomes are clustered on two health outcomes, a simple switch of the frequencies of these two health outcomes leaves the index of health inequality unchanged.
- *Invariance to parallel shifts*: When all individual health outcomes are clustered on two health outcomes, a parallel shift of the entire frequency distribution leaves the index of health inequality unchanged.
- *Polarization*: A "median preserving" change in the spread of a frequency distribution increases its inequality. A simple illustration is a move from the distribution {0, 0.3, 0.4, 0.3, 0} to the distribution {0.3, 0, 0.4, 0, 0.3}.

The focus of Yalonetzky's (2016) note is on the paper published by Lv et al. (2015). The latter authors propose a class of measures that, on one hand cardinalize the distances between ordinal categories, but on the other hand fulfill key properties, like aversion to median preserving spreads, that are desirable when dealing with ordinal variables. Of particular interest is the fact that the indices proposed by Lv et al. (2015) have a property which Yalonetzky labels Kolm-independence which guarantees that the change in total inequality due to a change in the relative frequency of an ordinal category is independent of the initial level of that frequency. However, since Lv et al. (2015) derive several measures that have this so-called Kolm-independence property, Yalonetzky wonders how robust pairwise ordinal inequality comparisons are to alternative choices of equally appropriate ordinal inequality measures. He then derives the first-order stochastic dominance condition that will guarantee that all inequality measures belonging to the Kolm-independent class will rank a pair of distributions consistently.

Cowell and Flachaire (2017) start by stressing the fact that there are essentially two ways in which the literature on inequality measurement dealt with ordinal data. One possibility is to first impute a notion of "status" to a categorical data structure, then to examine the inequality of status. Such an imputation may be derived from some subjective evaluation by individuals, via, for example, a Likert scale. But such an approach implies the use of a cardinalization which is arbitrary. A second solution is to focus on first-order dominance criteria. With ordinal data the median plays then the role of the mean in traditional inequality 015) analysis. Note however that, as stressed by Abul Naga and Yalcin (2010), comparing distributions with different medians may be problematic. Cowell and Flachaire propose therefore a different approach which includes three main elements: the notion of status within a distribution, a reference point and a set of axioms. For them status can be downward- or upward-looking, depending on the context of the analysis. These authors characterize a family of indices that depends on a sensitivity parameter and a reference point. Their axiomatic derivation ends up with a specific family of inequality measures related to the generalized entropy and Atkinson classes. It is important to stress that the reference point for categorical data is not the mean of the distribution but either the maximum or minimum possible value of the status.

More precisely Cowell and Flachaire (2017) start by stating what they call "mergers principle." Assume we have a classification of health in four categories: "bad health," "fair health," "good health," and "very good health." Add now a category called "very bad health" but suppose that no one is in this category. Cowell and Flachaire (2017) then state that we can then ignore this empty category and merge it with the next category below or above. They also assume that if two categories are merged, this should not have any impact on any individual not classified in these two categories. However if we now consider the case where the category to which individual i belongs is merged with an adjacent category, the impact of this merger on the status of individual i will depend on whether we assume a "downward-looking" or an "upward-looking" status.

Let f_i refers to the proportion of individuals who are in category i and call F_i the cumulative distribution, that is, the proportion of individuals who are in category i or in a lower category. Then the "peer-inclusive downward-looking" status s_i of an individual belonging to category i will be expressed as $s_i = F_i$. The "peer-inclusive upward-looking" status S_i of an individual belonging to category i will be expressed as $S_i = \sum_{k=i}^{I} f_k$ so that $S_1 = 1$ and $S_I = f_I$. In what follows we will use the notations relevant for the case of "peer-inclusive downward-looking" status.

Call e the reference point to which the elements of the status vector s will be compared. To look at inequality, the authors define a distance function $d(s, e)$ which refers to the distance between an individual with status s and the reference point e. In addition call $s(\zeta, i)$, the vector obtained when the i^{th} component of s is replaced by ζ. They then define principles allowing them to characterize an inequality ordering $\geq$ and the corresponding distance concept.

Here are the axioms that Cowell and Flachaire list:

- *Axiom 1*: *Continuity*. $\succcurlyeq$ is continuous.
- *Axiom 2*: *Monotonicity in distance*. Assume two status vectors s and s' that differ only in their i^{th} element.
  If $s_i^{'} \geq e$, then $s_i > s_i^{'}$ implies that $(s, e) > \left(s', e\right)$.
  If $s_i^{'} \leq e$, then $s_i < s_i^{'}$ implies that $(s, e) > \left(s', e\right)$.
- *Axiom 3*: *Independence*. If $s(\zeta, i)) \sim s'(\zeta, i)$ for some ζ, then $s(\zeta, i)) \sim s'(\zeta, i)$ for any ζ.
- *Axiom 4: Anonymity.* Any permutation $\prod$(s) of s is such that $\left(\prod(s), e\right) \sim (s, e)$.

  The first theorem they derive on the basis of these axioms establishes inequality as the total distance from the reference point.

  In a second stage, Cowell and Flachaire add the following axioms:
- *Axiom 5*: *Scale invariance*. The inequality orderings (not the level of inequality) remain unchanged when status is rescaled.

They then prove that Axioms 1 to 5 lead to measures of the form (or strictly increasing functions of them) $I_\alpha(s, e)$ where

$$I_\alpha(s, e) = \frac{1}{\alpha(\alpha - 1)}\left\{\left[\left(\frac{1}{n}\sum_{i=1}^{n}(s_i)^\alpha\right) - e^\alpha\right]\right\} \tag{7.48}$$

Note that (7.48) implies that $I_\alpha(1e, e) = 0$, where $1e$ refers to a vector s where all elements are equal to e. Note also that the average status $s_{average}$ in the population is defined as $s_{average} = \sum_{i=1}^{I} f_i s_i$, and it is easy to derive that we can also write that $s_{average} = \sum_{i=1}^{I} f_i S_i$.

Cowell and Flachaire consider then various possible reference points: maximum status ($e = 1$), minimum status ($e = 0$), mean status $(e = s_{average})$ and median status. They reach the conclusions:

- Selecting as reference the median or mean status leads to unsatisfactory or strange results.
- If one takes a “peer-inclusive” definition of status, that the reference point should be the maximum status.
- If one takes a “peer-exclusive” definition of status, that the reference point should be the minimum status.

Note that in (7.48), the smaller α, the greater the weight given to small status, relative to high status, values.

In the particular case where $\alpha = 0$, Eq. (7.48) will be expressed as

$$I_0(s, e) = \log e - \left(\frac{1}{n}\sum_{i=1}^{n}\log\, s_i\right) \tag{7.49}$$

If $\alpha = 1$, Eq. (7.48) will be expressed as

$$I_1(s, e) = \left(\frac{1}{n}\sum_{i=1}^{n} s_i \log(s_i)\right) - e\log e,\ \text{if}\, e = s_{average}. \tag{7.50}$$

But if $e \neq s_{average}$, $I_1(s, e) = \pm\infty$.

Cowell and Flachaire stress however the fact that $I_\alpha(s, e)$ is well-behaved only if $\alpha < 1$.

Another possibility is to use Atkinson indices $A_\alpha(s)$ with

$$A_\alpha(s) = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}(s_i)^\alpha\right]^{1/\alpha} \text{ if } \alpha < 0 \text{ or } 0 < \alpha < 1. \tag{7.51}$$

$$A_\alpha(s) = 1 - \left[\prod_{i=1}^{n} s_i\right]^{1/n} \text{ if } \alpha = 0. \tag{7.52}$$

Empirical illustrations

These methods that are specifically designed for ordinal data are used in a number of studies on health status. Abul-Naga and Yalcin (2008) illustrate their method using self-assessed health from the 2002 wave of the Swiss Health Survey and show variations in inequality across the seven Swiss regions. They also highlight that the choice of parameters matters. Kobus and Milos (2012) re-use these data and employ their decomposition technique to compute the contribution of each of the seven Swiss regions to total inequality.

Pascual et al. (2018) compute the health polarization and inequality indices developed by Apouey (2007) and Abul-Naga and Yalcin (2008), using different parameter values. Their data come from two waves (2006–2009 and 2013–2015) of the European Health Interview Survey (EHIS) and they focus on self-assessed health (in five categories). Findings highlight the persistence of significant health inequality in a number of European countries. Moreover, particularly high levels of inequality are observed in Cyprus, Greece, the Czech Republic, and Hungary. In addition, Madden (2011) focuses on mental health and psychological well-being in Ireland between 1994 and 2001, i.e., a period that incorporates an economic boom, using data from the Living in Ireland Survey (LII). Taken together, his results show falling health inequality during that period of high economic growth. Indeed, using the S-dominance criterion (Allison & Foster, 2004), he observes decreasing inequality in some dimensions of satisfaction, in self-assessed health, and in the General Health Questionnaire stress score. Moreover, the Apouey (2007) and the Abul-Naga and Yalcin (2008) measures generally indicate a decline in polarization and inequality over time. Di Novi et al. (2019) also assess inequality in self-assessed health, but across Italian regions. Their findings based on the Kobus and Milos (2012) index shows that health inequalities are relatively high in Italy compared to other European countries. Interestingly, the analysis suggests that fiscal autonomy is associated with lower health inequalities in Italy. This result is supported by an econometric analysis.

Wang and Yu (2016) examine the distribution of self-assessed health in China. Their data come from the China Health and Nutrition Survey (CHNS) between 1997 and 2009, and the authors compute the Apouey (2007) index for different values of α, among other indicators. The analysis depicts a rather striking evolution of health inequality, in line with the strengthening of income inequality in the country. Indeed, using the calibrated value of α, health inequality continuously increased over time, and this increase is very strong and reaches 100%. A strengthening of inequality is also found when urban and rural areas are studied separately. Inequalities in urban areas are greater than inequalities in urban areas.

Methods for ordinal data have not only been employed to study the distribution of individual health status, but also that of opinions regarding health systems. In particular, Jones et al. (2011) use measures developed by Abul-Naga and Yalcin (2008) and Apouey (2007) to study inequality and polarization in the responsiveness of health systems in Europe. While health systems should have three goals—population health, fairness of financing, and responsiveness—according to the World Health Report (World Health Organization), the authors argue that inequality in the responsiveness of health systems has received little attention. To bridge this gap, the authors analyze individual-level data from the World Health Survey on 25 European countries. Their four outcomes of interest capture individual ratings on "clarity of communication," "dignity," "confidentiality," and "prompt attention" in the process of care, with response categories running from "very good" to "very bad." Findings highlight the existence of inequality and polarization in responsiveness. Moreover, substantial variability in inequality and polarization is found across countries.

Conclusion

The first studies on health inequality were unidimensional and applied indices such as the Gini index to health data (Le Grand, 1987). Methodological research then moved to the measurement of bidimensional (i.e., social) health inequality (O'Donnell et al., 2007; Wagstaff &

Van Doorslaer, 2004). This approach has become very popular and the concentration index is now a widely used indicator in the health economics literature. An important feature of the Gini coefficient and of the concentration index is that they may be decomposed into factors, to unravel the causes of overall and social health inequalities. Recent research has developed tools to capture health inequality for ordinal health variables (Abul-Naga and Yalcin, 2008; Apouey, 2007; Lazar & Silber, 2013). Note that indicators for ordinal data are relevant not only for health and health care outcomes, but also in other domains, such as happiness and life satisfaction (see, e.g., Dutta, 2013; Madden, 2011).

In this chapter, we reviewed a number of these indicators of univariate and bivariate inequality and presented empirical illustrations highlighting the usefulness of these approaches. The empirical literature shows large differences in health (e.g., life expectancy) between poor and rich countries. Moreover, research highlights the existence of significant levels of overall and social health inequalities within countries. In other words, the burden of poor health is greater among poorer social groups throughout the world. Important factors derived from a decomposition analysis of the indicators could serve as targets to improve equality. Inequality levels and the contributions of explanatory factors differ between countries.

The measurement of overall and social health inequalities will require further research. First, most papers have focused on some specific health and health care measures (such as self-assessed health, body weight, mortality, or out-of-pocket payments). However, health inequalities tools could be applied more broadly to other health scores (such as mental health scores and clinical measures). In addition, for certain health outcomes, different results may be reached when using different inequality indices. In that case, additional thought about the choice of the indices will be needed.

We conclude with two comments to improve the description and understanding of health distributions in a population. First, health poverty and richness indicators (Simões et al., 2016) may be a nice complement to health inequality measures, in detailed descriptions of health distributions. Second, while bivariate inequality indices highlight the link between social position and health, we cannot infer a causal interpretation from these indicators and their decomposition. However, we believe that knowledge about causal paths is necessary to design policies. The use of additional techniques (exogenous shocks, instrumental variables, etc.) is thus an important complement to the descriptive approach, based on indices, that was reviewed in this chapter.

References

Abul Naga, R. H., & Yalcin, T. (2008). Inequality measurement for ordered response health data. *Journal of Health Economics, 27*(6), 1614–1625.

Abul Naga, R. H., & Yalcin, T. (2010). *Median independent inequality orderings*. SIRE Discussion Papers 2010–118, Scottish Institute for Research in Economics (SIRE). Aberdeen, UK.

Adesanya, O. A., Darboe, A., Mendez Rojas, B., Abiodun, D. E., & Beogo, I. (2017). Factors contributing to regional inequalities in acute respiratory infections symptoms among under-five children in Nigeria: A decomposition analysis. *International Journal for Equity in Health, 16*, 140.

Adeyanju, O., Tubeuf, S., & Ensor, T. (2017). Socio-economic inequalities in access to maternal and child healthcare in Nigeria: Changes over time and decomposition analysis. *Health Policy and Planning, 32*(8), 1111–1118.

Allison, R. A., & Foster, J. E. (2004). Measuring health inequality using qualitative data. *Journal of Health Economics, 23*(3), 505–524.

Apouey, B. H. (2007). Measuring health polarization with self-assessed health data. *Health Economics, 16*, 875–894.

Apouey, B. H. (2010). On measuring and explaining socioeconomic polarization in health with an application to French data. *Review of Income and Wealth, 56*(1), 141–170.

Apouey, B. H. (2015). Les disparités sociales de santé perçue au cours de la vie: Le cas de la France (2004–2012). *Bulletin Epidémiologique Hebdomadaire, 24–25*, 456–465.

Apouey, B. H., & Silber, J. (2016). Performance and inequality in health: A comparison of child and maternal health across Asia. *Research on economic inequality* (Vol. 24). *Inequality after the 20th century: Papers from the sixth ECINEQ meeting* (pp. 181–214). Emerald Group Publishing Limited.

Ataguba, J. E., Akazili, J., & McIntyre, D. (2011). Socioeconomic-related health inequality in South Africa: Evidence from general household surveys. *International Journal for Equity in Health, 10*, 48.

Ataguba, J. E.-O., Day, C., & McIntyre, D. (2016). Explaining the role of the social determinants of health on health inequality in South Africa. *Global Health Action, 9*(s3), 28865.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2*, 244–263.

Bado, A. R., & Sathiya Susuman, A. (2016). Women's education and health inequalities in under-five mortality in selected Sub-Saharan African countries, 1990–2015. *PLoS ONE, 11*(7), e0159186.

Balia, S., & Jones, A. M. (2008). Mortality, lifestyle and socio-economic status. *Journal of Health Economics, 27*, 1–26.

Cowell, F. A., & Flachaire, E. (2017). Inequality with ordinal data. *Economica, 84*(334), 290–321.

Deaton, A. (2013). What does the empirical evidence tell us about the injustice of health inequalities? In N. Eyal, S. Hurst, O. F. Norheim, & D. Wikler (Eds.), *Inequalities in health: Concepts, measures and ethics*. Oxford University Press.

Di Novi, C., Piacenza, M., Robone, S., & Turati, G. (2019). Does fiscal decentralization affect regional disparities in health? Quasi-experimental evidence from Italy. *Regional Science and Urban Economics, 78*, 103465.

Duclos, J.-Y., Esteban, J., & Ray, D. (2004). Polarization: Concepts, measurement, estimation. *Econometrica, 72*(6), 1737–1772.

Dutta, I. (2013). Inequality of happiness in the U.S.: 1972–2010. *Review of Income and Wealth, 59*(3), 393–415.

Erreygers, G. (2009). Correcting the concentration index. *Journal of Health Economics, 28*(2), 504–515.

Erreygers, G., & Van Ourti, T. (2011a). Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice. *Journal of Health Economics, 30*(4), 685–694.

Erreygers, G., & Van Ourti, T. (2011b). Putting the cart before the horse. A comment on Wagstaff on inequality measurement in the presence of binary variables. *Health Economics, 20*, 1161–1165.

Esteban, J.-M., & Ray, D. (1994). On the measurement of polarization. *Econometrica, 62*(4), 819–851.

Hicks, N., & Streeten, P. (1979). Indicators of development: The search for a basic needs yardstick. *World Development, 7*(6), 567–580.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior, 38*(1), 21–37.

Ikilezi, G., Augusto, O. J., Sbarra, A., Sherr, K., Dieleman, J. L., & Lim, S. S. (2020). Determinants of geographical inequalities for DTP3 vaccine coverage in Sub-Saharan Africa. *Vaccine, 38*, 3447–3454.

Jones, A. M., Rice, N., Robone, S., & Rosa Dias, P. (2011). Inequality and polarisation in health system's responsiveness: A cross-country analysis. *Journal of Health Economics, 30*, 616–625.

Jürges, H. (2010). Health inequalities by education, income and wealth: A comparison of 11 European countries and the US. *Applied Economics Letters, 17*(1), 87–91.

Kivimäki, M., Batty, G. D., Pentti, J., Shipley, M. J., Sipilä, P. N., Nyberg, S. T., Suominen, S. B., Oksanen, T., Stenholm, S., Virtanen, M., Marmot, M. G., Singh-Manoux, A., Brunner, E. J., Lindbohm, J. V., Ferrie, J. E., & Vahtera,

J. (2020). Association between socioeconomic status and the development of mental and physical health conditions in adulthood: A multi-cohort study. *Lancet Public Health, 5*(3), e140–e149.

Kjellsson, G., & Gerdtham, U. G. (2013). On correcting the concentration index for binary variables. *Journal of Health Economics, 32*(3), 659–670.

Kobus, M., & Milos, P. (2012). Inequality decomposition by population subgroups for ordinal data. *Journal of Health Economics, 31*, 15–21.

Kobus, M., Polchlopek, O., & Yalonetzky, G. (2019). Inequality and welfare in quality of life among OECD countries: Non-parametric treatment of ordinal data. *Social Indicators Research, 143*, 201–232.

Kolm, S. C. (1969). The optimal production of social justice. In J. Margolis & H. Guitton (Eds.), *Public economics*. Macmillan.

Lazar, A., & Silber, J. (2013). On the cardinal measurement of health inequality when only ordinal information is available on individual health status. *Health Economics, 22*, 106–113.

Le Grand, J. (1987). Inequalities in health. Some international comparisons. *European Economic Review, 31*, 182–191.

Lv, G., Wang, Y., & Xu, Y. (2015). On a new class of measures for health inequality based on ordinal data. *Journal of Economic Inequality, 13*, 465–477.

Madden, D. (2011). The impact of an economic boom on the level and distribution of subjective well-Being: Ireland, 1994–2001. *Journal of Happiness Studies, 12*, 667–679.

Mané, P. Y. B. (2013). Decomposing health care use inequalities in Senegal. *d'Economie du Développement, 21*(1), 61–89.

Marmot, M., & Allen, J. (2020). COVID-19: Exposing and amplifying inequalities. *Journal of Epidemiology and Community Health, 74*, 681–682.

Mendelson, H. (1987). Quantile-preserving spread. *Journal of Economic Theory, 42*, 334–351.

Nkonki, L. L., Chopra, M., Doherty, T. M., Jackson, D., & Robberstad, B. (2011). Explaining household socio-economic related child health inequalities using multiple methods in three diverse settings in South Africa. *International Journal for Equity in Health, 10*, 13.

O'Donnell O., Van Doorslaer, E., & Van Ourti, T. (2015). "Health and inequality," Chapter 17. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of income distribution* (Vol. 2B). North Holland.

O'Donnell, O., Van Doorslaer, E., Wagstaff, A., & Lindelow, M. A. (2007). *Analyzing health equity using household survey data. A guide to techniques and their implementation*. WBI Learning Resources Series, The World Bank, Washington.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review, 14*(3), 693–709.

Paes de Barros, R., Ferreira, F. H. G., Molinas Vega, J. R., & Saavedra Chanduvi, J. (2008). *Measuring inequalities of opportunities in Latin America and the Carribean*. The World Bank.

Paes de Barros, R., Molinas Vega, J. R., & Saavedra Chanduvi, J. (2008). Measuring inequality of opportunities for children. Background paper for *Measuring inequality of opportunities in Latin America and the Carribean*. The World Bank.

Pascual, M., Cantarero, D., & Lanza, P. (2018). Health polarization and inequalities across Europe: An empirical approach. *The European Journal of Health Economics, 19*, 1039–1051.

Quentin, W., Abosede, O., Aka, J., Akweongo, P., Dinard, K., Ezeh, A., Ramadan, H., Kalambayi Kayembe, P., Mitike, G., Mtei, G., Te Bonle, M., & Sundmacher, L. (2014). Inequalities in child mortality in ten major African Cities. *BMC Medicine, 12*, 95.

Reardon, S. F. (2009). Measures of ordinal segregation. In *Research on Economic Inequality* (Vol. 17, pp. 129–155). Emerald.

Sen, A. K. (1973). *On economic inequality*. Clarendon.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica, 50*(197), 3–17.

Silber, J. (1983). ELL (The equivalent length of life) or another attempt at measuring development. *World Development, 11*(1), 21–29.

Silber, J. (1989). Factors components, population subgroups and the computation of the gini index of inequality. *The Review of Economics and Statistics, 71*(1), 107–115.

Simões, N., Crespo, N., Moreira, S. B., & Varum, C. A. (2016). Measurement and determinants of health poverty and richness: Evidence from Portugal. *Empirical Economics, 50*, 1331–1358.

Theil, H. (1967). *Economics and information theory*. North-Holland.

Tranvåg, E. J., Ali, M., & Norheim, O. F. (2013). Health inequalities in Ethiopia: Modeling inequalities in length of life within and between population groups. *International Journal for Equity in Health, 12*, 52.

Van Doorslaer, E., & Jones, A. M. (2003). Inequalities in self-reported health: Validation of a new approach to measurement. *Journal of Health Economics, 22*, 61–87.

Van Malderen, C., Amouzou, A., Barros, A. J. D., Masquelier, B., Van Oyen, H., & Speybroeck, N. (2019). Socioeconomic factors contributing to under-five mortality in sub-Saharan Africa: A decomposition analysis. *BMC Public Health, 19*, 760.

Van Malderen, C., Van Oyen, H., & Speybroeck, N. (2013). Contributing determinants of overall and wealth-related inequality in under-5 mortality in 13 African countries. *Journal of Epidemiology and Community Health, 67*(8), 667.

Wagstaff, A. (2002). Inequality aversion, health inequalities and health achievement. *Journal of Health Economics, 21*(4), 627–641.
Wagstaff, A. (2005). The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics, 14*, 429–432.
Wagstaff, A. (2011a). The concentration index of binary outcome revisited. *Health Economics, 20*, 1155–1160.
Wagstaff, A. (2011b). Reply to Guido Erreygers and Tom Van Ourti's comment on "The concentration of a binary outcome revisited." *Health Economics, 20*, 1166–1168.
Wagstaff, A., & Van Doorslaer, E. (2004). Overall versus socioeconomic health inequality: A measurement framework and two empirical illustrations. *Health Economics, 13*, 297–301.
Wagstaff, A., Van Doorslaer, E., & Watanabe, N. (2003). On decomposing the causes of health sector inequalities, with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics, 112*, 219–227.
Wang, H., & Yu, Y. (2016). Increasing health inequality in China: An empirical study with ordinal data. *Journal of Economic Inequality, 14*, 41–61.
Wolfson, M. C. (1994). When inequalities diverge. *American Economic Review, 84*(2), 353–358.
Yalonetzky, G. (2016). *Robust ordinal inequality comparisons with Kolm-independent measures* (ECINEQ Working Papers 401).

CHAPTER 8

Inequality of Opportunity: Theoretical Considerations and Recent Empirical Evidence

Flaviana Palmisano and Vito Peragine

Introduction

The notion of "equal opportunities" has been of long-standing relevance in public debates and is increasingly proposed as a principle of social justice by politicians of different orientations. As an example, goal 10 of the United Nation's Sustainable Development Goals (SDG) recognizes that "income inequality cannot be effectively tackled unless the underlying inequality of opportunities is addressed".

However, the meaning of equality of opportunity remains often vague in the public discourse, and this may partly explain its popularity. Though

F. Palmisano
Sapienza University of Rome, Rome, Italy
e-mail: flaviana.palmisano@uniroma1.it

V. Peragine (✉)
University of Bari, Bari, Italy
e-mail: vitorocco.peragine@uniba.it

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_8

multiple definitions exist, the essence of the concept is that equality of opportunity is obtained when everyone exerting the same degree of effort (or responsibility) attains the same level of advantage (or well-being), regardless of any predetermined circumstances beyond their control (Roemer, 1998). Outcome inequalities are therefore consistent with equal opportunities only to the extent that they derive from differences in factors individuals can be held responsible for.

There are different reasons for embracing the opportunity perspective. The first is that most of those who worry about inequality do so because they think that it is unjust, or at least partially unjust. In addition, existing surveys show that most people judge income inequalities arising from different levels of effort as less objectionable than those due to exogenous circumstances as gender, race, family origin, etc. The implicit idea is that what matters for a just society is the distribution of opportunities, rather than the distribution of outcomes. Hence, it is interesting to measure that portion of outcome inequality that can be attributed to exogenous circumstances and that, thus, reflects unequal opportunities.

Along the same lines, prominent political philosophers propose a distinction between fair (justifiable) and unfair (unjustifiable) inequalities (Arneson, 1989; Cohen, 1989; Dworkin, 1981). According to these theories, inequality arising from factors over which the individual does not have any control—such as race, sex, ethnicity, religion, birthplace and family background—should be of primary concern from an ethical standpoint and should therefore be considered as unfair. On the other hand, inequality resulting from factors for which one can arguably be held responsible are regarded as fair.

In addition to normative reasons, the analysis of opportunity inequality can have an instrumental value. First, social attitudes towards redistributive policies may be affected by the knowledge, or the perception, of the origin of income inequalities (Alesina & La Ferrara, 2005). By recognizing that a small (large) amount of existing inequalities is due to unequal opportunities, one may decrease (increase) the support for redistributive policies. Second, opportunity inequality, rather than income inequality, can be related to aggregate economic performance: it has been suggested (Bourguignon et al., 2007a, 2007b; World Bank, 2006) that the existence of strong and persistent inequalities in the initial opportunities open to individuals can generate true inequality traps that represent severe constraints to perspectives of future growth of an economy, by preventing entire groups from participation into economic and social life.[1]

Finally, the analysis of opportunity inequality may help the understanding of the generation of income inequality since it constitutes the hardest layer to remove through public intervention. The knowledge of the factors determining opportunity inequality can help to identify the more deprived groups in a society, thereby revealing new points of emphasis in social and redistributive policies.

Mainly inspired by the philosophical debate on responsibility-sensitive egalitarian justice, Roemer (1993, 1998), Van de Gaer (1993) and Fleurbaey (1995, 2008) have proposed formal economic models in which inequality of opportunity (IOp) is defined as the part of overall inequality that is generated by factors beyond an individual's control. Following these seminal contributions, a rich literature has flourished in the past two decades, proposing different approaches and methodologies to measure the degree of inequality of opportunity in different dimensions of well-being, time periods and countries (see Ferreira & Peragine, 2016; Roemer & Trannoy, 2016; Ramos & Van de Gaer, 2016). The diversity in methodological approaches has offered a variety of empirical evidence.

This chapter aims at proposing a critical discussion of this rich literature, focusing mainly on the empirical results offered by existing studies.

It is articulated as follows. Section "The Equality of Opportunity Approach" introduces the canonical theoretical model of equality of opportunity (section "The Theoretical Model") and an empirical model which has been extensively used in the literature (section "The Empirical Model"). Section "Empirical Evidence" offers a review of the recent most relevant empirical findings on inequality of opportunity. It first discusses studies that focus on developed countries (section "Inequality of Opportunity in Developed Countries"); then, it presents studies that cover less developed countries (section "Less Developed Countries"). Last (section "The Global Perspective"), by adopting a global perspective, it discusses empirical evidence based on a recent dataset

[1] For an empirical analysis of the relationship between inequality of opportunity and growth in a sample of US states see Marrero and Rodríguez (2013); they decompose total inequality into inequality of opportunity and inequality of effort, showing that GDP per capita growth rate is negatively correlated with the former and positively with the latter. A similar line of research has been followed by Ferreira et al. (2014), with a cross-country analysis involving a sample of 84 countries.

(the EqualChances.org database), created in order to obtain consistent comparisons of inequality of opportunity across countries and outcomes. Section "Conclusions" concludes.

The Equality of Opportunity Approach

The Theoretical Model

Consider a distribution of outcome x in a given population. Suppose that all determinants of x, including the different forms of luck, can be classified into either a set of circumstances C that lie beyond individual control, or as responsibility characteristics, summarized by a variable[2] e, denoting effort. Circumstances belong to a finite set Ω. For example, suppose that the only circumstance variables are race that can only take values in the set {black, white}, and parental education that only takes values in the set {college education, high school education}. In this case the set Ω would be the following: Ω = ({black, parents with high school education}, {black, parents with college education}, {white, parents with high school education}, {white, parents with college education}).

Effort may be treated as either a continuous or a discrete variable belonging to the set Θ. The outcome of interest is generated by a function $g : \Omega \times \Theta \rightarrow R$ such that:

$$x = g(C, e) \tag{8.1}$$

This is a reduced-form model in which outcomes are exclusively determined by circumstances and effort, such that all individuals having the same circumstances and the same effort obtain the same outcome. Neither opportunities themselves, nor the process by which some outcomes are chosen, are explicitly modelled in this framework. The idea is to infer the opportunities available to individuals by observing joint distributions of circumstances, effort, and outcomes. Roughly speaking, the source of unfairness in this model is given by the effect that circumstance variables (which lie beyond individual responsibility) have on individual outcomes.

Thus, there is a population of individuals, each of whom is fully characterized by the triple (x, C, e). For simplicity, treat effort e, as well as each element of the vector of circumstances, C, as discrete variables.

[2] Effort could also be treated as a vector. However, following the literature, it is treated as a scalar in this chapter.

Table 8.1 Distribution of outcomes according to circumstances and effort

	e_1	e_2	e_3	...	e_m
C_1	x_{11}	x_{12}	x_{13}	...	x_{1m}
C_2	x_{21}	x_{22}	x_{23}	...	x_{2m}
C_3	x_{31}	x_{32}	x_{33}	...	x_{3m}
...	...	...	...	...	...
C_n	x_{n1}	x_{n2}	x_{n3}	...	x_{nm}

Then this population can be partitioned in two ways: into types T_i, within which all individuals share the same circumstances, and into tranches T_j within which everyone shares the same degree of effort. Denote by x_{ij} the outcome generated by circumstances C_i and effort e_j. Suppose, in addition, that there are n types, indexed by $i = 1, \ldots, n$, and m tranches, indexed by $j = 1, \ldots, m$. In this discrete setting,[3] the population can be represented by a matrix $\left[X_{\text{ij}}\right]$ with n rows, corresponding to types, and m columns, corresponding to tranches.

To the $n \times m$ dimensional matrix $\left[\mathbf{X}_{\text{ij}}\right]$ in Table 8.1, let there be associated a $n \times m$ dimensional matrix $\left[\mathbf{P}_{\text{ij}}\right]$ where each element p_{ij} represents the proportion of total population with circumstances C_{i} and effort e_{j}.

Given this model, the measurement of inequality of opportunity can be thought of as a two-step procedure: first, the actual distribution $\left[\mathbf{X}_{\text{ij}}\right]$ is transformed into a counterfactual distribution $\left[\tilde{\mathbf{X}}_{\text{ij}}\right]$ that reflects only and fully the unfair inequality in $\left[\mathbf{X}_{\text{ij}}\right]$, while all the fair inequality is removed. In the second step, a measure of inequality is applied to $\left[\tilde{\mathbf{X}}_{\text{ij}}\right]$. The construction of the counterfactual distribution $\left[\tilde{\mathbf{X}}_{\text{ij}}\right]$ should reflect the principle of equality of opportunity.

Within this framework, the opportunity egalitarian principle can be decomposed into two distinct and independent sub-principles: *the Reward Principle*, which is concerned with the apportion of outcome to effort and, in some of its formulations, requires to respect the outcome inequalities due to effort; and *the Compensation Principle*, according to

[3] In an alternative formulation, that would treat effort as a continuous variable, $F_i(x)$ would denote the advantage distribution in type i and q_i denote its population share. The overall distribution for the population as a whole would be $F(x) = \sum_{i=1}^{n} q_i F_i(x)$.

which all outcome inequalities due to C are unfair and should be compensated by society. Any satisfactory measure of opportunity inequality should respect both the compensation and the reward principles.

The existing literature has developed two main versions of the compensation principle and two consequent approaches to the measurement of opportunity inequality, namely the ex-ante and the ex-post approach.

According to the ex-ante approach, there is equality of opportunity if the set of opportunities is the same for all individuals, regardless of their circumstances. Hence in the ex-ante version, the compensation principle is formulated with respect to individual opportunity sets: it requires reducing the inequality between opportunity sets (*ex-ante compensation*). In the model introduced above, a given row i, that is the outcome distribution of a given type, is interpreted as the opportunity set of all individuals with circumstances C_i. Hence, the focus is on the rows of the matrix above: the counterfactual distribution should reflect the inequality between the rows.

On the other side, according to the ex-post approach, there is equality of opportunity if and only if all those who exert the same effort end up with the same outcome. The compensation principle, in the ex-post version, is thus defined with respect to individuals with the same effort but different outcomes: it requires reducing outcome inequality among the individuals with the same effort (*ex-post compensation*). This means that opportunity inequality within this approach is measured as inequality within the columns of the matrix. Hence, the corresponding counterfactual distribution should reflect the inequality within the columns.

As far as the *reward* principle is concerned, different versions of the principle have been proposed by the literature, expressing different attitudes with respect to the outcome inequality observed among individuals endowed with the same circumstances: from utilitarian reward (Fleurbaey, 2008; Van de Gaer 1993) which expresses perfect neutrality, to inequality averse reward (Ramos & Van de Gaer, 2016) which expresses aversion to inequality, to intermediate and agnostic positions (Fleurbaey & Peragine, 2013; Peragine, 2002).

Different measures, which are either consistent with the ex-ante or the ex-post approaches, and with different versions of reward, have been proposed in the literature (see Ferreira & Peragine, 2016; Ramos & Van de Gaer, 2016): they express different and sometimes conflicting views on equality of opportunity and in fact the rankings they generate may

Table 8.2 Measuring between-types inequality $(n = m = 3)$

	e_1	e_2	e_3
C_1	μ_1	μ_1	μ_1
C_2	μ_2	μ_2	μ_2
C_3	μ_3	μ_3	μ_3

be different.[4] In addition, their informational requirements are quite different: while for the ex-ante approach one needs to observe the individual outcome and the set of circumstances, for the ex-post approach a measure of individual effort is required. Therefore, in addition to normative considerations, the choice of the methodology to adopt should also reflect the data availability. As often the database does not contain a satisfactory measure of effort, most of empirical applications focus on the ex-ante approach.

A measure extensively used in the literature, based on ex-ante compensation and utilitarian reward, is *Between-Types* inequality proposed in its non-parametric version by Peragine (2002) and Checchi and Peragine (2010). It relies on a counterfactual distribution $\left[\tilde{X}_{\mathrm{BT}}\right]$ that is obtained replacing each individual outcome x_{ij} by the average outcome of the type she belongs to (μ_i), abstracting from her level of effort (see Table 8.2). This smoothing transformation is intended to remove all inequality within types. Formally:

$$\boldsymbol{Between-types\ counterfactual\ distribution}$$

$$\left[\tilde{\mathbf{X}}_{\mathrm{BT}}\right]: \forall j \in \{1, \ldots, m\}, \forall i \in \{1, \ldots, n\}, \tilde{x}_{\mathrm{ij}} = \mu_i = \frac{\sum_{j=1}^{m} p_{\mathrm{ij}} x_{\mathrm{ij}}}{\sum_{j=1}^{m} p_{\mathrm{ij}}}$$

It is immediate to notice that between-types inequality is consistent with the principle of utilitarian reward: the types of $\tilde{X}_{\mathrm{BT}}$ are made up of replications of the same outcome, the mean, and therefore the artificial distribution does not reflect any inequality within type—the kind of inequality which is fair according to the reward principle, and thus

[4] See Fleurbaey and Peragine (2013) for a discussion of the clash between ex-ante and ex-post equality of opportunity.

should be cleansed in $\left[\tilde{\mathbf{X}}_{\mathrm{ij}}\right]$. It is also consistent with ex-ante compensation, as the inequality between types (evaluated as the inequality between the means of each type) is preserved. Once the smoothed distribution $\left[\tilde{\mathbf{X}}_{\mathrm{BT}}\right]$ is obtained, any inequality measure I applied to such distribution $I\left(\tilde{\mathbf{X}}_{\mathrm{BT}}\right)$ is to be interpreted as a measure of inequality of opportunity.

An alternative, ex-post measure, inspired by Roemer's (1993) and implemented by Checchi and Peragine (2010) and Aaberge et al. (2011), is based on the *Within Tranches* counterfactual distribution ($\tilde{X}_{\mathrm{WTR}}$). It is obtained by replacing each individual outcome x_{ij} in a given tranche with the ratio between such outcome and the average outcome of that tranche: $\boldsymbol{v_j} = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{p_{\mathrm{ij}} x_{\mathrm{ij}}}$. This normalization procedure is intended to remove all inequalities between tranches and to leave unchanged the inequality within tranches. Formally:

$$\text{Within tranches}\left(\tilde{X}_{\mathrm{WTR}}\right): \text{For all } j \in \{1, \ldots, m\} \text{ and}$$
$$\text{For all } i \in \{1, \ldots, n\}, \tilde{x}_{ij} = g(c_i, e_j)/v_j.$$

It is easy to see that within tranches is consistent with ex-post compensation: each tranche is obtained simply by rescaling original outcomes by a constant $(1/v_j)$. Therefore, $\tilde{X}_{\mathrm{WTR}}$ accounts for all the original (relative) inequality within tranches. On the other hand, compliance with the reward principle is not guaranteed, since Table 8.3 does in general contain inequality within types: for at least one i and a couple j, h, $\tilde{x} = g(c_i, e_j)/v_j \neq g(c_i, e_h)/v^h = \tilde{x}_{\mathrm{ih}}$.

Once the counterfactual distribution has been obtained, either in the ex-ante or in the ex-post versions, the specific inequality index *I(.)* does vary across different papers as it will be discussed in the next section.

Table 8.3 Within tranches inequality ($n = m = 3$)

	e1	*e2*	*e3*
C1	x_{11}/v_1	x_{12}/v_2	x_{13}/v_3
C2	x_{21}/v_1	x_{22}/v_2	x_{23}/v_3
C3	x_{31}/v_1	x_{32}/v_2	x_{33}/v_3

The Empirical Model

The ex-ante between-types measure $I(\tilde{x}_{\mathrm{BT}})$ has been extensively implemented in empirical analyses of inequality of opportunity by a number of authors.

All these papers use a measure of economic well-being—mostly household per capita income, household per capita consumption, or individual labour earnings—as the advantage indicator. For this reason, Brunori et al. (2013) refer to the between-types measure of IOp in these studies as an index of Inequality of Economic Opportunity (IEO). Two closely related versions of the index are often reported: the absolute or level estimate of inequality of opportunity ($\mathrm{IEO_L}$), given simply by the inequality measure computed over $\tilde{X}_{\mathrm{BT}}$, i.e. by $I(\tilde{x}_{\mathrm{BT}})$. The ratio of $\mathrm{IEO_L}$ to overall inequality in the relevant advantage variable (e.g. household per capita income), which yields the relative measure, $\mathrm{IEO_R}$:

$$\mathrm{IEO}_R = \frac{I(\tilde{\mathrm{x}}_{\mathrm{BT}})}{I(\mathrm{x})} \tag{8.2}$$

The partition of types varies across studies (see Brunori et al., 2013). Because in some cases, the data sets are not large enough to yield precise estimates of $\boldsymbol{\mu_i}$ for all types, some authors compute $\mathrm{IEO_L}$ using a parametric approximation. After estimating the reduced-form regression of income on circumstances:

$$\mathrm{x} = \mathrm{C}\beta + \in \tag{8.3}$$

and obtaining coefficient estimates $\hat{\beta}$, these authors use predicted incomes as a parametric approximation to the smoothed distribution:

$$\boldsymbol{I}\left(\hat{\boldsymbol{x}}_{\mathbf{BT}}\right), \text{where } \hat{\boldsymbol{x}}_{\boldsymbol{i}} = \boldsymbol{C}_{\boldsymbol{i}}\hat{\boldsymbol{\beta}} \tag{8.4}$$

Parametric estimates are also presented either as levels ($\mathrm{IEO_L}$) or ratios ($\mathrm{IEO_R}$), analogously. This approach follows Ferreira and Gignoux (2011), which in turn draws on Bourguignon et al. (2007a, 2007b). This method is particularly useful when the number of circumstances to be included in the analysis (which usually depends on data availability) gets larger. Under this condition, the non-parametric approach would be based on a partition of the population into types containing a small number of individuals and hence would face the risk of higher estimation bias.

It is important to note that these empirical estimates of "between-types" IOp—whether estimated parametrically or non-parametrically—are, in each and every case, lower-bound estimates of inequality of opportunity. A formal proof of the lower-bound result is contained in Ferreira and Gignoux (2011) but the intuition is straight-forward: the set of circumstances which is observed empirically—and used for partitioning the population into types—is a strict subset of the set of all circumstance variables that matter in reality. The existence of unobserved circumstances—virtually a certainty in all practical applications—guarantees that these estimates of IOp could only be higher if more circumstance variables were observed.

As far as the inequality index is concerned, several papers, following the pioneering work by Checchi and Peragine (2010), have used the Mean Log Deviation (MLD), which is a member of the entropy family of inequality measures, well-known for its decomposability property. In fact, given a population and a partition into homogeneous sub-groups, the MLD is perfectly decomposable into a between- and within-group component:

$$\mathrm{MLD} = \mathrm{MLD_W} + \mathrm{MLD_B}$$

In the context of inequality of opportunity, by defining the groups as sets of individuals sharing the same circumstances, the "within-group" term is interpreted as inequality due to effort and the "between-group" term is interpreted as inequality due to circumstances, i.e. inequality of opportunity (see Checchi & Peragine, 2010).

An alternative measure, increasingly used in the context of inequality of opportunity, is the Gini coefficient. Its use is justified by several arguments. First, the MLD is very sensitive to extreme values, much more than the Gini coefficient. Its high sensitivity implies that the reduction of inequality generated by transforming the original into the smoothed distribution (the first step of our procedure) will be much higher for the MLD than for the Gini coefficient. Incidentally, *ceteris paribus*, relative IOp as measured by the MLD will be much lower than relative IOp as measured by the Gini coefficient. Symmetrically, the MLD is insensitive to small levels of inequality that typically characterize between-group inequality with sufficiently large groups. By using MLD we would therefore obtain estimates levelled towards zero and this would limit our ability to appreciate between-country difference in IOp.

On the other hand, the Gini coefficient is not strictly decomposable into a within and a between-groups component. Instead, a decomposition of Gini into between- and within-group inequality leaves a positive residual term whenever the supports of the distributions for different groups overlap. In general:

$$G = G_W + G_B + K$$

where K is a residual, greater than zero when groups' distributions overlap. In the context of inequality of opportunity, K measures the part of inequality that is jointly determined by effort and circumstances, but that cannot be disentangled into the effect of effort and of circumstances. Researchers then have tried to understand how much of a problem this positive residual poses in measuring inequality of opportunity. The answer might depend on the interpretation of the within-group inequality component of the decomposition. In the case that econometricians consider that they have included every relevant circumstance variable in the partition of the population (so that all material circumstance variables are observed and used to construct the counterfactual distribution), the within-group component of the decomposition could be safely interpreted as inequality due to effort.

However, in the more likely empirical setting where not all relevant circumstances are observed, the within-group component should be—and usually it is—treated as residual. The partial observability of circumstances is known to generate downward bias in the measurement of inequality of opportunity, and empirical estimates of IOp are usually interpreted as lower-bound measures of IOp. In this case, the fact that the residual term of the Gini decomposition is always positive means that the sum $G_W + K$ can be treated as the residual term. Therefore, G_B can be interpreted as a lower-bound measure of IOp, and inequality of opportunity measured through the Gini index can be expressed in both its absolute and relative version, as follows:

- Absolute inequality of opportunity: abs-IOp = $\text{Gini}(\tilde{Y})$
- Relative inequality of opportunity: rel-IOp = $\text{Gini}(\tilde{Y})/(\text{Gini}(\text{Ywithin}) + \text{Gini}(\tilde{Y}) + K)$

Note that, as far as the relative IOp index is concerned, K is part of the denominator but not part of the numerator. This makes rel-IOp more conservative than if the normalization were obtained with a perfectly decomposable inequality measure.

The choice between MLD and Gini in IOp analysis is still debated among specialists and a comparison of the features inherent to each index does not lead to an unambiguous superiority of one index over the other.

Although the predominance of the Gini coefficient and the Mean Logarithmic Deviation in empirical analyses, the Dissimilarity index has also become a common choice for the measurement of inequality of opportunity when the outcome variable is binary. Another set of studies, instead, following the theoretical insights proposed by Peragine (2002, 2004) and Lefranc et al. (2009), has explored the possibility of performing IOp comparisons across distributions employing partial rankings criteria.

Empirical Evidence

This section discusses some of the recent empirical findings on inequality of opportunity. To give a systematic order to this review, a distinction will be made between studies that focus on developed countries and studies that cover less developed countries. Then, for each subgroup of countries, the discussion will concern inequality of opportunity measured in the space of monetary outcome (income, earnings, or consumption). Since the principle of equality of opportunities has also been extended in the empirical literature for considering other non-monetary outcomes with a meaningful economic interpretation, this chapter will also review empirical findings concerning inequality of opportunity measured in the non-monetary space and, more specifically, in the education and health space.

The reader must be aware that different studies make different choices in terms of: sample selection,[5] specification of the outcomes,[6]

[5] For instance, whether to focus only on men or only on women or on both; age/cohort range of the sample, place of residence (urban, rural, or both), etc.

[6] In the case of monetary outcome, the researcher needs to choose, for instance, among: equivalized household disposable income, individual income, per capita consumption, individual labour income, etc. In the case of education outcome, the researcher needs to choose, for instance, among: years of schooling, graduation marks, highest education attainment, access to tertiary education, etc. In the case of health outcome, the researcher needs to choose, for instance, among: body mass income, self-perceived health status, child nutrition, etc.

list of circumstances,[7] specification of the effort variable,[8] measurement methodology implemented.[9] There are no contributions that operationalize the inequality of opportunity concept making the same choices for treating different data. At the same time, with some exceptions that will be discussed at length in section "The Global Perspective", there are no contributions that apply all the different methodological options to the same dataset. Therefore, the reader should always use a certain degree of caution when comparing the results provided by the existing empirical works—as well as those provided by the subset of works that will be surveyed in the next sections.

For this reason, this chapter will mostly review studies that focus on groups of countries using harmonized data, rather than studies that estimate IOp based on country-specific surveys, which would further hamper cross-countries comparability.

Table 8.4 offers a summary prospect of the contributions that will be discussed in the next pages distinguished by country group (developed countries, less developed countries, global) and outcome variable (monetary, non-monetary and by type of non-monetary variables, namely education and health).

Inequality of Opportunity in Developed Countries

Inequality of Opportunity in the Monetary Space

Empirical research on IOp in the monetary space in developed countries using harmonized data is mainly based on data from European countries and the United States.

As regards European countries, most studies rely on the European Union–Statistics on Income and Living Conditions database (EU-SILC)

[7] In general, all observed and meaningful circumstances are included in the analysis since the higher the number of circumstances included the lower is the estimation bias due to unobserved exogenous factors.

[8] Observable information of effort is rarely available. Hence, the choice here becomes whether to use the Roemer's identification axiom stating that those that are at the same percentile of the distribution of income conditional on their type have exercised the same degree of effort.

[9] Ex-ante or ex-post approach; parametric or non-parametric estimation; choice among a variety of inequality indicators; choice between complete or partial rankings, etc.

Table 8.4 Summary prospect of the studies reviewed

Geographic areas	*Outcome*		*Papers reviewed*
Developed countries	Monetary		Checchi et al. (2016) Ramos and Van de gaer (2020) Brunori et al. (2022) Bussolo et al. (2019) Marrero and Rodriguez (2011)
	Non-monetary	Education	Lasso de la Vega et al. (2020) Palmisano et al. (2022)
		Health	Bricard et al. (2013) Brunori et al. (2021)
Less developed countries	Monetary		Brunori et al. (2019) Singh (2012) Choudhary et al. (2019) Brock et al. (2017) Alvarez and Menendez (2020)
	Non-monetary	Education	Gamboa and Waltenberg (2015) Brock et al. (2017)
		Health	Aizawa (2019) Perez-mesa et al. (2020)
Global	Monetary		Equalchances.org (2018) Milanovic (2015)

for obtaining comparable estimates on inequality of opportunity. EU-SILC collects comparable information on socioeconomic and demographic characteristics of individuals across European countries. Particularly relevant for researchers interested in inequality of opportunity are the 2005 and 2011 waves, since they provide information on family background and circumstances when the respondent was young.[10] The main limitation of EU-SILC is the reduced sample sizes for some countries,

[10] The 2005 wave consists of the following countries: Austria (AT), Belgium (BE), Cyprus (CY), Czech Republic (CZ), Germany (DE), Denmark (DK), Estonia (EE), Greece (EL), Spain (ES), Finland (FI), France (FR), Hungary (HU), Ireland (IE), Iceland (IS), Italy (IT), Lithuania (LT), Luxemburg (LU), Latvia (LV), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Sweden (SE), Slovenia (SI), Slovakia (SK) and Great Britain (UK). In the 2011, Bulgaria (BG), Switzerland (CH), Croatia (HR), Malta (MT) and Romania (RO) are added to the previous list, counting 31 countries in total.

which obliges to work with a reduced number of circumstances (and efforts).

Using EU-SILC data in 2005, a variety of contributions find that IOp for income is lowest in Nordic countries and Slovenia, by contrast and not surprisingly, Mediterranean countries (Italy, Greece, Spain and Portugal), Anglo-Saxons (Greta Britain and Ireland) and poorer Eastern EU (Estonia, Latvia, Poland and Lithuania) are characterized by the highest inequality of opportunity (see, among others, Checchi et al., 2010; Marrero & Rodriguez, 2012).

More recently, Checchi et al. (2016) use EU-SILC 2005 and 2011 but, differently from previous contributions that used equivalized household disposable income, are interested in understanding ex-ante IOp for individual incomes and labour market positions. Indeed, they argue that family members may be characterized by different sets of circumstances (gender and age for instance vary across family members); if this is the case, then, averaging among members would attenuate the impact of individual circumstances. In addition, mating, family formation and fertility are individual choices, which according to the IOp theory should be kept separate from circumstances. The circumstances included in their analysis are gender, age, country of origin and family background and the sample is restricted to individuals aged between 30 and 60 who are either working full or part-time, unemployed or fulfilling domestic tasks and care responsibilities. Their non-parametric estimates of the Gini IOp index uncover a particular picture: the usual country ranking—Nordic countries performing better according to most distributional phenomena and Mediterranean countries performing worst—does not hold when the evaluation metric is represented by IOp. Furthermore, IOp across European countries appears to be more persistent than total inequality.

To provide a methodologically robust analysis on IOp on EU-SILC data, Ramos and Van de Gaer (2020) report that choosing between ex-ante and ex-post approaches is not immune to sizeable influence on country ranking. Moreover, this is the methodological choice that matters more in terms of ranking differentials across alternative measurement models. These findings confirm and corroborate previous theoretical results proving the incompatibility between ex-ante and ex-post approaches to inequality of opportunity (see, among others, Fleurbaey & Peragine, 2013). The second most relevant methodological choice that arises from their study is whether one adopts a norm-based measure or other measures, thus recognizing that, as for the case of mobility, inequality of opportunity is a multifaceted concept.

The need to provide more robust evaluations of IOp using EU-SILC data also motivates Brunori et al. (2022) work. They investigate the possibility of introducing machine learning methods in the empirical analysis of IOp, and their results seem encouraging. They prove that the adoption of machine learning tools reduces the risk of ad-hoc model selection and that selecting estimation models without following a non-arbitrary rule may lead to an overestimation (underestimation) of IOp by up to 30% (40%) in comparison to the method that is generated by the machine learning algorithm.

A different database to estimate IOp in European countries is used by Bussolo et al. (2019): this is the Luxembourg Income Study database (LIS), which provides microdata on a large number of countries and time. The data are originally collected by country-specific institutions (usually central banks or national statistical institutes) and harmonized by the LIS team. The countries considered in this comparative study are Italy, Germany, France, Switzerland and United Kingdom over the last two decades. The outcome variable is disposable income and is harmonized according to the LIS procedures. Income is then converted to constant prices using the national consumer price index. The list of circumstances used is given by: gender, age, parental background, country of origin. The sample is composed of individuals aged 25 to 80. Making these choices, Bussolo et al. (2019) parametrically estimate ex-ante IOp using the MLD and find that there is a great variability across the countries considered. In some instances (Italy and Germany) inequality of opportunity represents an important portion of total income inequality, with values ranging from 20 to 50% but it has been smoothly declining over the period considered. In some other cases (France, UK and especially Switzerland), this portion is more limited, reaching at most 35%. A more in-depth analysis indicates that IOp decreases with age implying that the role played by circumstances at birth becomes weaker over the life cycle. This represents a remarkable distinguishing feature between IOp and standard income or consumption inequality, which is often characterized by an increasing trend. By contrast, the cohort effect is mixed: a decreasing trend of IOp is revealed for UK and Germany, with the younger generation experiencing lower IOp levels; an inverted U-shape path is instead found for Italy and France.

The United States represent another region of the developed world that has been largely investigated in terms of inequality of opportunity. The database most used for this purpose is the Panel Study of Income Dynamics (PSID), a household panel representative of the entire US population, started in 1968 and still running. The PSID contains information on individual income and on more than one circumstance and it is highly accurate. Marrero and Rodriguez (2011) use this dataset to measure ex-ante IOp in United States and consider individuals who are household heads aged between 25 and 50. The outcome variable is gross income computed as the household head's labour income plus the household capital income divided by the number of adults in the household. The circumstances considered are father's education and race. They (parametrically and non-parametrically) estimate a declining trend for inequality of opportunity (measured through the MLD) in the United States between 1969 and 1977 that becomes constant and at low levels between 1977 and 1985. IOp starts to increase between 1986 and 1997, but from 1997 through 2005 the trend appears to be very much approach-dependent: parametrically estimated IOp considerably drops, whereas non-parametrically IOp remains relatively stable. The last period of their analysis is 2005–2007, during which a renewed drop in IOp is shown. The decomposition of total inequality (obtained applying the natural decomposition of the squared coefficient of variation and the Nested Shapley value) reveals that IOp accounts for less than 10% of total income inequality over the entire period. 5 to 20% is explained by a correlation between effort and circumstances. Race is the main circumstance during the 1970s and 1980s, accounting for more than 50%, overcome by parental education in the last two decades.[11]

Inequality of Opportunity in the Non-Monetary Space: Education

IOp in education (EIOp) is not only normatively relevant, but it is also positively relevant. Indeed, the distribution of educational achievements may affect the distribution of earnings, as predicted by the human capital theory, and the potential for growth.

Most of the empirical contributions aimed at evaluating IOp in education use the relationship between children's circumstances and inequalities in standardized test scores measured in international surveys, especially

[11] Marrero and Rodriguez (2011) represent an update of the results provided in previous works by Pisoltesi (2009).

the PISA (Programme for International Student Assessment).[12] PISA is collected every 3 years since 2000 and includes data on test scores of representative samples of students aged 15 in dozens of countries in a large number of countries on three subjects—mathematics, sciences and reading—as well as detailed information on students' background and schools' personnel and functioning conditions. Using such data sources allows for consistent cross-country comparisons as they provide standardized measures of achievements and the same set of information at individual and school levels (see, among others, Betts & Roemer, 2005; Ferreira & Gignoux, 2010, 2014; Gamboa & Waltenberg, 2015; Salehi-Isfahani et al., 2014).

Lasso de la Vega et al. (2020) use the 2012 wave of PISA to estimate educational IOp in 20 selected countries.[13] Their list of circumstances includes gender, family background, school background and peer group effect. They opt for a parametric estimation of IOp that includes observable measures of effort and to capture the direct effect of circumstances on overall inequality. They find that, among the countries selected, effort contributes to generating total educational inequality more than circumstances in Finland, Iceland, and Norway. Romania, Finland, Spain, Ireland, Greece and Norway show not only the lowest levels of total educational inequality but also the lowest levels of EIOp, while Belgium arises to be the worst performing country.

Less explored due to data availability is inequality of opportunity for higher education. Most of the existing works are country-specific (Brunori et al., 2012; Jaoul-Grammare & Magdalou, 2017; Peragine & Serlenga, 2008). An exception is Palmisano et al. (2022) who use the EU-SILC database for two survey years, 2005 and 2011 to assess IOp for tertiary education. Their outcome is a binary variable indicating whether an individual has achieved tertiary education or not; their sample is composed of working-age individuals (individuals aged 25 to 60). The circumstances used are parental education, parental occupation, area of birth, gender and financial problems when the individual was a teenager.

[12] Other surveys used for the same purposes are TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study).

[13] Belgium, Bulgaria, Croatia, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Lithuania, Luxembourg, Netherlands, Norway, Portugal, Romania, Spain, Sweden, Switzerland.

Their results—obtained through a parametric estimation of the dissimilarity index to measure educational inequality of opportunity and robust to the period considered—reveal that, although there exist some relevant across-country variations, Mediterranean and Eastern European countries not only perform badly in the space of income but also in the space of education, while EIOp is lowest for Northern European countries. Parental education and occupation arise to be the most relevant circumstances, pointing to socioeconomic background as the most relevant driver of educational IOp. They argue that policies aimed at improving equality of opportunity in higher education to be effective should focus on programmes informing students and their families of the benefits of tertiary education, should consider the introduction or reinforcement of need-based grants and should be directed at reducing the burden of tertiary education on the students or on their family of origin.

Inequality of Opportunity in the Non-monetary Space: Health
Turning to health, the other non-monetary outcome considered in this survey, empirical research on inequality of opportunity is mostly based on data from single European countries for adult populations (see, among others, Li Donni et al., 2014, 2015; Trannoy et al., 2010; Van de Gaer et al., 2014). Nevertheless, it is possible to report at least two contributions that extend the analysis to more than one country using harmonized data.

Bricard et al. (2013) use data from the Retrospective Survey of SHARELIFE, the third wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) collected in 2008/2009, since it focuses on life histories of European people aged 50 and over and thus provides a unique set of information on circumstances and health status for several European countries.[14] Their sample is composed of individuals aged between 50 and 80. Health is measure through a binary variable indicating whether or not the individuals rate their health as "good" or less than "good". The vector of circumstances includes several social conditions in childhood, parents' longevity and parents' health-related behaviours. Their parametric estimates of IOp in health reveal that in Europe about 50%

[14] SHARE is a representative dataset of the European population aged 50 and over in Scandinavia (Denmark and Sweden), Western Europe (Austria, France, Germany, Switzerland, Belgium and the Netherlands), Mediterranean countries (Spain, Italy and Greece), and two transition countries (the Czech Republic and Poland).

of total health inequality is explained by health inequality of opportunities, but there are large variations across countries. Austria, France, Spain and Germany are characterized by higher level of health IOp. Whereas Sweden, Poland, Belgium, the Netherlands and Switzerland present the lowest levels. Social and family determinism of lifestyles play the most relevant role in determining IOp in health in Belgium, the Netherlands, Italy, Germany, Poland and Denmark. Hence, for these countries, the authors argue, it would be desirable to reducing social reproduction and the intergenerational transmission of unhealthy lifestyles. In Austria, France, Spain and Czech Republic, instead, high IOp in health is mainly driven by social and family background. For these country policies compensating for poorer initial conditions would be beneficial.

Estimating health IOp in Europe is also the main aim of Brunori et al. (2021), who use the EU-SILC database (2011), considering the sample of individuals aged 25 to 60. Their binary outcome variable indicates whether or not the individuals rate their health as "at least fair", while the set of circumstances includes gender and variables describing the socioeconomic background of all respondents. As regards the measurement method, they opt for the sequential dominance criterion to rank countries based on the latent class approach proposed by Li Donni et al. (2015) and Carrieri et al. (2020). The latent type approach is based on the idea that if types provide a sensible partition of the population, within-type inequality due to circumstances should be minimal. By contrast, between-type inequality due to circumstances should be maximum. Latent types are identified so to minimize within-type homogeneity in terms of circumstances, that is, maximizing between-type heterogeneity in terms of circumstances. They then introduce a*n opportunity-inequality curve*, allowing for a ranking of distributions that is robust to the number of types included in the analysis. They show that Eastern Europe countries are bottom ranked while Mediterranean countries, such as Spain, Malta, Greece and Cyprus, are among the best performers. A more in-depth analysis allows them to infer that inequality of health opportunity is strongly and negatively correlated with age-adjusted self-reported health conditions.

Less Developed Countries

Inequality of Opportunity in the Monetary Space

The work by Brunori et al. (2019) represents the first attempt to evaluate inequality of opportunity in a large set of SSA countries. They focus on Comoros, Ghana, Guinea, Madagascar, Malawi, Niger, Nigeria, Rwanda, Tanzania and Uganda during the 2000s. Their analysis is based on a sub-sample of the original data obtained by considering only individuals aged 15 or older. The outcome variable chosen is per capita consumption. Although different surveys are used, the results of their analysis are comparable across countries since the consumption variable is adjusted for inflation and translated into 2011 purchasing power parity (PPP) international dollars. The list of circumstances includes gender and variables capturing socioeconomic background, such as, ethnicity. The result of their analysis uncovers a dramatic picture: exogenous circumstances account between 40 and 56% of total inequality for the generality of countries considered. This is a striking result, particularly if one considers that the computed measures are only lower-bound estimates of the inequality of opportunity level in each country. They also estimate the contribution of each single circumstance and show that Comoros, Ghana, Guinea and Niger are characterized by a large impact of birthplace. Father education is, instead, the most important circumstance in Madagascar, Malawi, Rwanda and Tanzania. It is mother education for Nigeria and ethnicity in Uganda. Their results differ substantially from the only previous contribution that has focused on inequality of opportunity in SSA. Cogneau and Mesplé-Somps (2008) analysed five SSA countries (Ivory Coast, Ghana, Guinea, Madagascar and Uganda) between 1985 and 1994. They use a very coarse set of circumstances (parental background) and, in fact, their results show a much lower level of inequality of opportunity than the one reported in Brunori et al. (2019). With some variation between countries, Cogneau and Mesplé-Somps's (2008) estimates show that the portion of inequality attributed to exogenous circumstances is between 10 and 20%. Brunori et al. (2019) estimates are, instead, in line with those made available by Shimeles and Nabassaga (2018) who focus on IOp to understanding why total inequality is so high in Africa. Shimeles and Nabassaga (2018) find that IOp is one of the dominant components of inequality (the others being ethnic fractionalization, limited tertiary education, and poor governance), counting for about the 30% of overall inequality (see also Brunori et al., 2018; Fosu, 2018).

Singh (2012) estimates inequality of opportunity in household per capita earning and household per capita consumption expenditure for men aged 21–65 in rural and urban India. Exploiting data from the India Human Development Survey (IHDS), Round I (2004–2005), which provides data on parental background factors, caste and region of birth used in this analysis as circumstance variables, the opportunity share of inequality in earning is found to vary between 18 and 26% for urban areas, and between 16 and 21% for rural India. He also concludes that parental education specific opportunity share of overall inequality in earning as well as in consumption expenditure is largest in urban India. For rural India, along with parental education, caste and geographical region also seem to play an important role in the overall inequality of opportunity estimates.

Choudhary et al. (2019) explore wave 2011–2012 of the IHDS to investigate on IOp among Indian Women. They include parental education, caste, religion and region of birth as circumstances, and income and consumption expenditure as economic outcomes. They show that overall IOp in income ranges from 18 to 25% (of total income inequality among women) in urban areas and from 16 to 21% in rural areas. The corresponding figures for consumption expenditure are 16–22% and 20–23% in urban and rural areas, respectively. Moreover, their findings show that parental education is the highest contributor to overall IOp in urban areas; it is region of birth in rural areas, followed by parental education and caste. Parental education not being the highest contributor to IOp in rural areas might be due to the fact that the lack of infrastructure in rural areas curtails the options available to parents related to schooling decisions about their children. Hence, higher parental education might not result in better education for their children which in turn will not convert into superior income. Religion seems also to play a substantial role for determining IOp in both rural and urban areas.

Alvarez and Menendez (2020) use LIS microdata, a database incorporating personal harmonized variables allowing for cross-country comparisons, not only for developed countries as discussed above, but also for developing countries. Their analysis considers Brazil, Egypt, Guatemala, India, Peru, and South Africa over the period from 2004 to 2014 and employs as outcome variable the disposable equivalized income (additional estimates for labour personal income, consumption and monetary consumption are also provided). Using as circumstances some demographic and parental background's factors, the authors find that South

Africa registers the highest level of IOp in 2008: 65% of total inequality is represented by IOp, but decreases afterwards reaching the 57%. India by contrast has levels of IOp that are not particularly high in 2004 (48%) and remain constant in 2011 (50%). Guatemala (2006) and Peru (2004) are also harmed by very high levels of IOp, 65 and 51%, respectively. However, these two countries achieve a great reduction in the indices analysed, reaching 44% in Guatemala in 2014 and 34% in Peru in 2013. An intermediate case is Brazil, with levels of IOp ranging from 0.541 to 0.463. Last, Egypt seems to perform better than the other countries considered, being 18% the share of total inequality due to circumstances. In all countries and for all periods, parental education arises to be the most important circumstance, accounting for more than half of IOp. These results depict two different time-trends for income inequality and IOp, with some countries improving significantly both kinds of inequality (Peru and Guatemala). India, instead, experiences a significant increase only in the levels of overall inequality. A small reduction in IOp, with income inequality keeping constant, is registered in South Africa, and a small decrease in both income inequality and IOp in Brazil. Their results in part corroborate previous findings (Bourguignon et al., 2007a, 2007b for Brazil, Belhaj Hassine, 2012 for Egypt, Singh, 2012 and Choudhary et al., 2019 for India); in other cases, comparisons cannot be executed due to the different methodological choices (for instance, Ferreira & Gignoux, 2011 for Brazil, Guatemala and Peru, Piraino, 2015 for South Africa).

IOp for income in Transition countries is scrutinized in the 2016–2017 EBRD Transition Report by Brock et al. (2017), exploiting the third round of the Life in Transition Survey (LiTS III) conducted by the World Bank and the EBRD in the second half of 2015 and the first half of 2016. Their sample is composed of working-age individuals from both rural and urban areas. The outcome variable selected is represented by individual self-reported income in the last 12 months; whereas the circumstances considered are gender, rural or urban place of birth, ethnicity, and factors related to parental background. The authors find that IOp is much greater for this group of countries compared to the western Europe but lower than in other emerging economies (e.g. Brazil and India). More than 30% of total income inequality is associated with IOp with however a large degree of variability across countries: IOp reaches high levels in a few transition countries that are now EU members (Bulgaria, Kosovo and Romania). Bosnia and Herzegovina, Montenegro and Serbia display some of the lowest estimates, comparable with those of Germany. Last, their results indicate that, in this region, there is a positive relationship between inequality of opportunity and income inequality, which becomes stronger among countries with higher inequality and weaker in countries with lower inequality.

Inequality of Opportunity in the Non-Monetary Space: Education
Gamboa and Waltenberg (2015) explore PISA 2006–2009 data to evaluate educational IOp in six Latin American countries (Argentina, Brazil, Chile, Colombia, Mexico and Uruguay) covering about the 70% of the total population in that region.[15] The outcome variable is the average score in each subject, while parental education, gender and school type are used as circumstances. Adopting a non-parametric estimation method, the authors find that educational IOp in Latin America ranges from less than 1% to up to 25%, depending on the year, the country, the subject, and the specification of circumstances. IOp seems to be mainly driven by parental education and school type. Brazil is characterized by the highest level of IOp in 2006 followed by Mexico; at the other extreme of the ranking one finds Argentina and Colombia. In 2009, Uruguay and Brazil are the most opportunity unequal countries, whereas Mexico and Chile are the leading countries.[16] Overall, their results call for the implementation of policies that would improve the initial learning conditions of more disadvantaged children—for instance, by improving the quantity and quality of pre-primary schools—or would compensate unfair differences—for instance, by providing additional teaching hours to pupils from disadvantaged types.

Brock et al. (2017), exploiting the same dataset and sample characteristics used to measure income IOp in transition countries, make light on IOp in tertiary education for the same group of countries. Differently from before, the outcome variable is represented by a binary variable indicating whether the individuals attained a tertiary education degree. Their study demonstrates that IOp in tertiary education, as measured by the dissimilarity index, is higher than IOp in income and is mostly determined by parental background, especially among the generation that came of age in the early 2000s than among the generation that entered

[15] See Ferreira and Gignoux (2010) for a similar analysis developed for the case of Turkey, finding that inequality of opportunity in the PISA standardised test scores for reading, maths and science accounts for between 27 and 33% of overall inequality.

[16] See also Golley and Kong (2018) for a recent analysis of educational IOp in China.

educational institutions before the subregion's transition to the market economy. Parents who achieved a tertiary education degree benefited more from the transition process than others. This is explained because of market liberalization that transformed tertiary education from being universally free to entailing significant costs. In countries where education is still free, scholarships awarded for covering the cost of living have effectively been gradually withdrawn. In these conditions, highly educated parents are clearly more likely to be in a better position to send their children to university and cover the relative costs. Place of birth (urban or rural area) is also an important factor impacting tertiary educational attainment. A less important role but still significant is, instead, played by parents' membership of the communist party. Overall, it is found that individual circumstances matter more for younger cohorts than for older cohorts.

Inequality of Opportunity in the Non-Monetary Space: Health

Exploiting the Demographic and Health Survey (DHS), Aizawa (2019) deals with the assessment of inequality of opportunity in health in 10 developing countries in Asia (Bangladesh, Nepal, Pakistan, Maldives, India, Cambodia, Myanmar, East Timor, Tajikistan, Kyrgyzstan). DHS gathers data collected by comparable nationally representative household surveys in more than 85 countries worldwide since 1984. The sample in the Aizawa's (2019) analysis encompasses children aged under five from both rural and urban areas. The outcome variable is a child's nutritional status, on whose determination the following circumstances are considered: factors related to parental health and socioeconomic condition, factors related to demographic and material living conditions. Given these choices, the highest degree of health IOp is found in Pakistan, the lowest in Maldives. In almost all countries, the high levels of health IOp are greatly determined by household affluence. Maternal health is a strong determinant of child nutrition in Bangladesh, Nepal, Pakistan, India, Cambodia, Myanmar, East Timor and Tajikistan. The difference in material standard of living is relevant in Pakistan, Maldives and India. Maternal exposure to the media is found to be relevant in Nepal, Pakistan, India and Myanmar. The difference in sanitary conditions exhibits a significant contribution in Nepal, India, Cambodia and East Timor. In addition, in most of the countries, the contribution of paternal education is sizably smaller than the one made by maternal education. This is not surprising given that in most Asian countries, mothers play a major role in raising children. This evidence suggests that policies giving priority to children from mothers with a less advantaged educational background should be promoted to enhance health IOp. Moreover, enhancing educational opportunities for children could lessen the inequality in nutritional status for future generations. More in general, priority should be given to children from disadvantaged households by promoting for instance conditional cash transfer programmes, which transfer money to more needy households conditional on investments in human capital.

The data contained in the DHS are also used in Perez-Mesa et al. (2020) to assess health IOp in 33 Sub-Saharan countries. They focus on children aged below 5 and as a health outcome use the standardized height-for-age z-score corrected by the age (in months) and gender. Information on family background, the mother socio-demographic and anthropometric factors, household structure, household facilities and the region of residence represent the circumstances in this analysis. They use parametric ex-ante estimation and apply the Gini index and the MLD. Their results indicate that child health inequality is systematically lesser for the cohort of 4–5 years old than for the younger cohorts. However, circumstances are impeding a further reduction in child health inequality. Indeed, the contribution of IOp to total health inequality has risen along with the age distribution in almost all countries, an evolution mostly explained by family background, household facilities and the place of residence.

The Global Perspective

As explained at the beginning of this section, comparability of results across different contributions is cumbersome and not always readily doable. Different authors use different data sources and, most importantly, make very different methodological choices.

Very recently the literature has made great efforts to improve comparability across countries, and it is with this aim that the EqualChances.org dataset has been created and launched in 2018. It is the first online repository of internationally comparable information on inequality of opportunity for 47 countries.[17] In so doing, it enables a global evaluation

[17] EqualChances.org also provides comparable estimated of income mobility across generations for 27 countries; measures of the intergenerational transmission of status for 41 countries; measures of educational mobility across generations for 148 countries.

of inequality of opportunity in the monetary space. The dataset provides absolute and relative measures of ex-ante IOp. Absolute IOp is measured through the Gini coefficient computed on the distribution of opportunities. Relative IOp is defined as the ratio between absolute inequality of opportunity and the total inequality in the distribution of the household equivalent disposable income measured through the Gini coefficient. To compute such measures three key circumstances are used for all countries, these are: parental education, parental occupation and origin, where origin may refer either to race or to ethnic origin or to parental culture or to parental religion or to area of birth. Countries for which information on one or more of these circumstances are unavailable are not included in the database.

Inequality of opportunity, measured in both absolute and relative terms, shows large variation across the world. In general, countries with high absolute level of IOp are also those in which the contribution of this inequality to total inequality is higher. On average, countries belonging to the American region feature higher level of IOp than European countries and Australia. Within each world region, emerging countries typically perform worse, compared to high-income economies (see Fig. 8.1).

Zooming on and comparing levels of absolute IOp, northern European countries are the best performer, with Iceland featuring the lowest value (0.029). At the other extreme of the ranking one finds African countries, with South Africa showing the highest value of IOp (0.337). Among high-income countries, the United States and Italy are those presenting the highest levels of absolute IOp, 0.137 and 0.148, respectively (see Fig. 8.2). This ranking is confirmed, when one considers relative IOp as the country ranking criterion, with relative IOp being highest in African and South American countries. Guatemala occupies the bottom rank (58%) and Denmark the top (12%) (see Fig. 8.3). Luxembourg and Italy are the developed countries showing the highest estimated levels of relative IOp, 0.451 and 0.457, respectively, that, except for Brazil, are considerably higher than relative IOp estimated for other European countries.

When more than two separate time-observations are available, it is possible to grasp some information on the evolution of IOp. Equalchance.org provides such observations for 8 countries (Australia, Brazil, Italy, Korea, Mexico, Panama, South Africa, United Kingdom). From Fig. 8.4, it comes out that the time trend is stable in all countries. There are a few exceptions to this: these are Panama and South

Fig. 8.1 Absolute and relative IOp around the world (*Notes* The panel on the left report absolute value. Estimates for each country are based on the most recent survey year available. *Source* The Equalchance.org database)

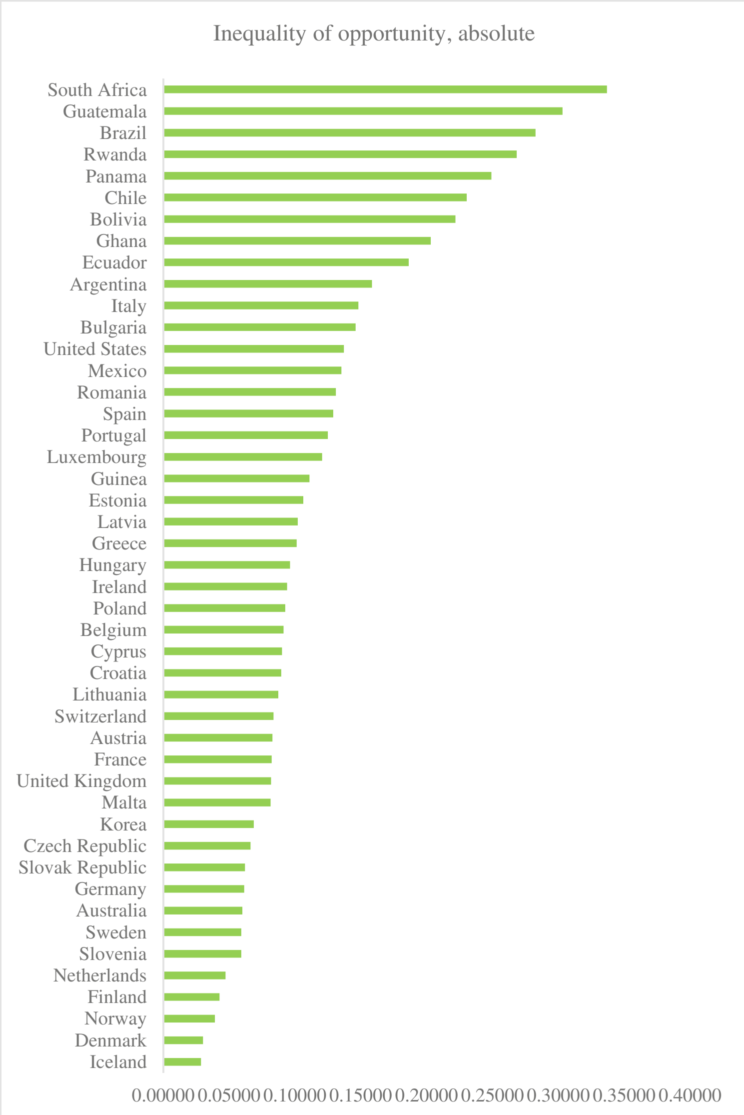

Fig. 8.2 Absolute IOp, country ranking (*Note* Estimates for each country are based on the most recent survey year available. *Source* The Equalchance.org database)

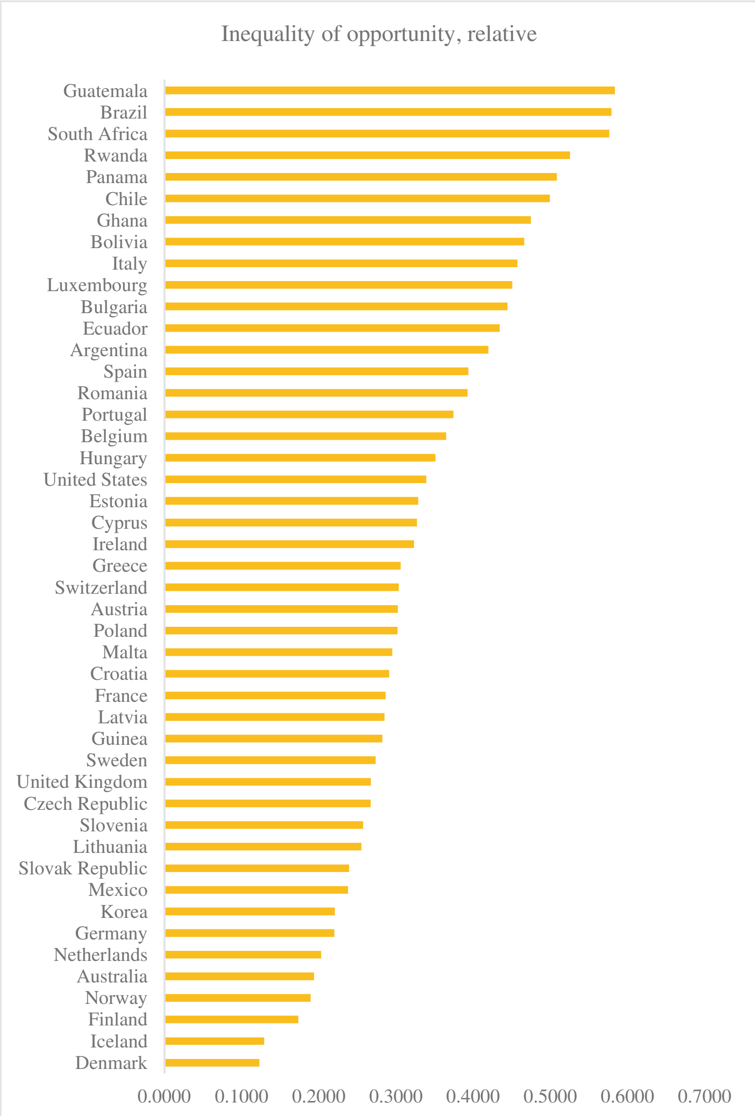

Fig. 8.3 Relative IOp, country ranking (*Note* Estimates for each country are based on the most recent survey year available. *Source* The Equalchance.org database)

Africa, experiencing a particularly decreasing trend. In some countries, the stability of the indexes is the result of opposing trends in the different sub-periods. This is the case of Brazil, a country in which IOp decreased during the 1990s and started to increase again afterwards. Differently in UK, IOp after an increase during the first half of 2000, went back to its initial values in 2010.

Milanovic (2015) also proposes a global perspective to the analysis of IOP. His global view is however different from the one adopted in the EqualChances.org study. In fact, while in the latter country boundaries are kept, in the former they are eliminated so to allow the analysis to be executed on the "world" population. To do this, Milanovic uses two circumstances: country of residence and income distribution within that country; he assumes that there is no migration and considers the household per capita income as the outcome variable. His estimates uncover that more than 50% of variability in income across the world citizens is accounted for by IOp. The contribution of the country of residence is substantial and disquieting: US citizens on average enjoy 350% of the per capita income enjoyed by an individual living in DR Congo—the poorest country in the world. For Brazilian citizens, the premium ranges around 160%, whereas it goes down to 32% for Yemen citizens.

Conclusions

The equality of opportunity literature has made important advances in the last twenty-five years, since the pioneering works of Roemer, Van de Gaer and Fleurbaey. Ideas first developed by philosophers, such as Arneson, Dworkin and Cohen, and by economists, such as Sen, have been translated into simple, coherent and powerful economic models by several economists since the 1990s, and a growing number of empirical applications have been proposed, with different methodologies and in different spheres of social life.

The existing methodological literature offers a solid guide for the estimation of inequality of opportunity in several different settings and socially relevant situations.

The equality of opportunity perspective can be used in distributional analysis to monitor social progress, as well as to evaluate the welfare effects of policy interventions in different areas. Combined with more standard approaches in welfare economics, such as the outcome-based approaches

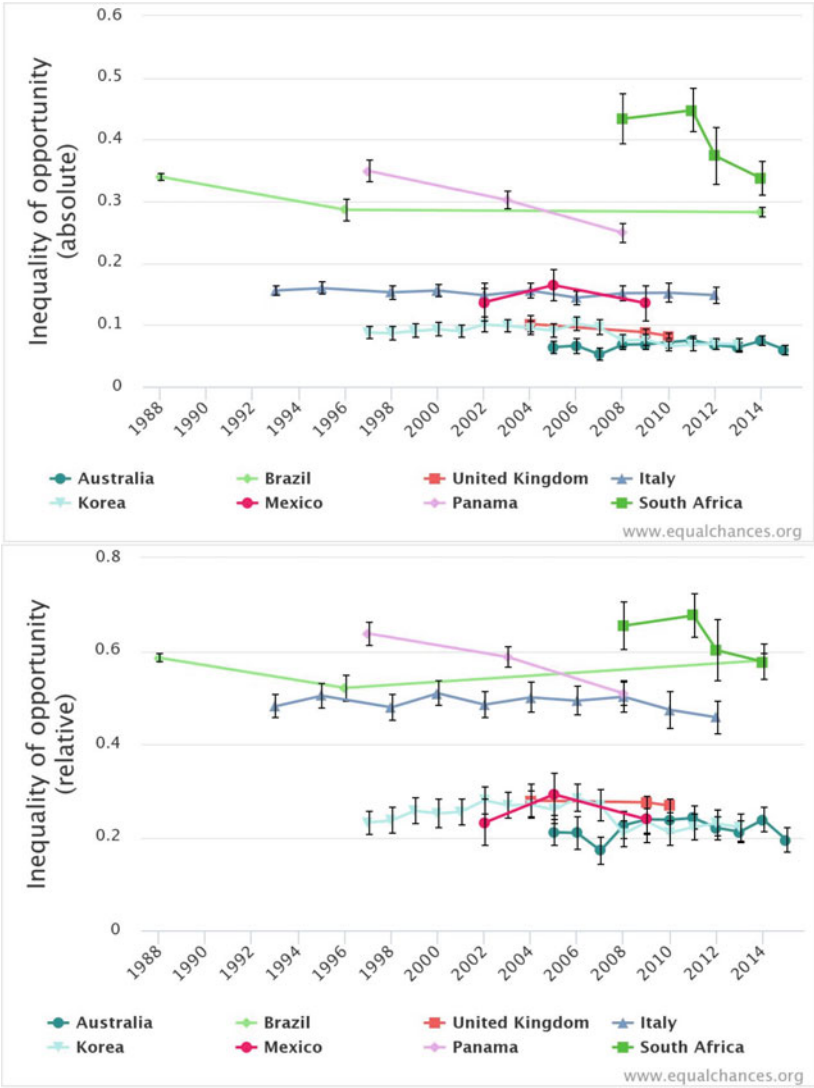

Fig. 8.4 Time trends of absolute and relative IOP (*Source* The Equalchances.org database)

to the evaluation of social inequities, it can help to identify emerging social needs and to set the priorities of redistributive policies.

Our review of the empirical literature shows that an important portion of economic inequality observed today in the world cannot be attributed to differences in the levels of individual efforts and responsibility. On the contrary, it can be directly ascribed to exogenous factors such as family background, gender, race, place of birth, etc. There is considerable cross-country variation in the (lower-bound) relative measure of ex-ante inequality. Although there certainly is noise in these measures, and various comparability caveats, there appears to be some signal as well.

In addition, the data collected show a positive correlation between inequality of opportunities and income inequality. Countries with a higher degree of income inequality are also characterized by greater inequality of opportunity. This result is consistent with the empirical literature on social mobility, which considers only one exogenous circumstance (family background measured on the basis of income or social status of the parents) and finds a negative correlation between inequality and mobility (see the "Great Gatsby Curve" of Corak, 2013): less unequal countries are also those that have a higher degree intergenerational mobility.

From the methodological viewpoint, while some important achievements have been reached and some consensus has started to emerge among specialists and practitioners, it is also true that the IOp literature is still young. A number of issues still need to be settled, which involve: (i) problems of comparability of inequality of opportunity estimates; (ii) robustness of IOp estimates to different methodological choices; (iii) potential extension of methods originally proposed for the measurement of inequality of opportunity for income, to alternative outcome spaces.

Two prominent issues arise in the empirical analysis of inequality of opportunity. The first is that of comparability across time and space: the diversity in methodological approaches and empirical specifications makes it difficult to create consistent comparisons of IOp across countries and outcomes. The recent construction of the global database EqualChances.org is an important step in this direction.

The second is that of the dimensions of inequality: while we have covered analyses of inequality of opportunity in different dimensions of social life (both monetary and non-monetary), they remain separate analyses that do not provide a true multidimensional assessment of inequality of opportunity. In this perspective, the recent theoretical models proposed by Kobus et al. (2020) could be an interesting path to follow.

These are (some of the) challenges for future works attempting at including a concern for responsibility and opportunities into the tools typically used to make normative evaluations and welfare comparisons.

References

Aaberge, R., Mogstad, M., & Peragine, V. (2011). Measuring long-term inequality of opportunity. *Journal of Public Economics, 95*(3–4), 193–204.

Aizawa, T. (2019). Ex-ante inequality of opportunity in child malnutrition: New evidence from ten developing countries in Asia. *Economics & Human Biology, 35*, 144–161.

Alesina, A., & La Ferrara, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics, 89*(5–6), 897–931.

Álvarez, A. S., & Menéndez, A. J. L. (2020). Trends in inequality of opportunity for developing countries: Does the economic indicator matter? *Social Indicators Research, 149*, 1–37.

Arneson, R. (1989). Equality of opportunity for welfare. *Philosophical Studies, 56*, 77–93.

Betts, J., & Roemer, J. E. (2005). *Equalizing opportunity for racial and socioeconomic groups in the United States through educational finance reform* (No. qt0gq4z4m9). Department of Economics, UC San Diego.

Belhaj Hassine, N. (2012). Inequality of opportunity in Egypt. *The World Bank Economic Review, 26*, 265–295.

Bourguignon, F., Ferreira, F. H. G., & Menéndez, M. (2007a). Inequality of opportunity in Brazil. *Review of Income and Wealth, 53*, 585–618.

Bourguignon, F., Ferreira, F. H. G., & Walton, M. (2007b). Equity, efficiency and inequality traps: A research agenda. *The Journal of Economic Inequality, 2*(5), 235–256.

Bricard, D., Jusot, F., Trannoy, L., & Tubeuf, S. (2013). Inequality of opportunities in health and the principle of natural reward: Evidence from European countries. *Research on Economic Inequality, 21*, 335–370.

Brock, M., Peragine, V., & Tonini, S. (2017). Inequality of opportunity. In Transition report 2016–2017. *Transition for all: Equal opportunities in an unequal world* (Chapter 3). EBRD.

Brunori, P., Ferreira, F. H. G., & Peragine, V. (2013). Inequality of opportunity, income inequality and mobility: Some international comparisons. In E. Paus (Ed.), *Getting development right: Structural transformation* (p. 2013). Palgrave Macmillan.

Brunori, P., Hufe, P., & Mahler, D. (2022). The roots of inequality: Estimating inequality of opportunity from regression trees and forest. Scandinavian Journal of Economics (forthcoming).

Brunori, P., Palmisano, F., & Peragine, V. (2018). Consumption dynamics and inequality of opportunity with an application to Uganda. *Review of Development Economics, 22*(2), 632–657.
Brunori, P., Palmisano, F., & Peragine, V. (2019). Inequality of opportunity in sub-Saharan Africa. *Applied Economics, 51*(60), 6428–6458.
Brunori, P., Peragine, V., & Serlenga, L. (2012). Fairness in education: The Italian university before and after the reform. *Economics of Education Review, 31*(5), 764–777.
Brunori, P., Trannoy, A., & Guidi, C. F. (2021). Ranking populations in terms of inequality of health opportunity: A flexible latent type approach. *Health Economics, 30*(2), 358–383.
Bussolo, M., Checchi, D., & Peragine, V. (2019). *Long-term evolution of inequality of opportunity* (World Bank Policy Research Working Paper N. 8700).
Carrieri, V., Davillas, A., & Jones, A. M. (2020). A latent class approach to inequity in health using biomarker data. *Health Economics, 29*(7), 808–826.
Checchi, D., Peragine, V., & Serlenga, L. (2016). Inequality of opportunity in Europe: Is there a role for institutions? In *Inequality: Causes and consequences*. Emerald Group Publishing Limited.
Checchi, D., Peragine, V., & Serlenga, L. (2010). *Fair and unfair income inequalities in Europe* (IZA Discussion Paper No. 5025/2010).
Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality, 8*(4), 429–450.
Choudhary, A., Muthukkumaran, G. T., & Singh, A. (2019). Inequality of opportunity in Indian women. *Social Indicators Research, 145*(1), 389–413.
Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics, 99*, 906–944.
Cogneau, D., & Mesplé-Somps, S. (2008). Inequality of opportunity for income in five countries of Africa. In J. Bishop and B. Zheng (Eds.), *Research on economic inequality* (Vol. 16, pp. 99–128). Emerald Group Publishing Limited.
Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives, 27*(3), 79–102.
Dworkin, R. (1981). What is equality? Part 1: Equality of welfare. Part 2: Equality of resources. *Philosophical Public Affairs, 10*, 185–246; 283–345.
Ferreira, F. H. G., & Gignoux, J. (2014). The Measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review, 28*(2), 210–246.
Ferreira, F. H. G., & Gignoux, J. (2010). Inequality of opportunity for education: Turkey. In R. Kanbur & M. Spence (Eds.), *Equity in a globalizing world, commission on growth and development* (pp. 131–156).

Ferreira, F. H. G., & Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth, 57*(4), 622–657.

Ferreira, F. H. G, & Peragine, V. (2016). Equality of opportunity: theory and evidence. In M. Adler & M. Fleurbaey (Eds.), *Handbook of well being and public policy*. Oxford University Press.

Fleurbaey, M. (1995). Three solutions for the compensation problem. *Journal of Economic Theory, 65*, 505–521.

Fleurbaey, M. (2008). *Fairness, responsibility and welfare* (1st ed.). Oxford University Press.

Fleurbaey, M., & Peragine, V. (2013). Ex-ante versus ex-post equality of opportunity. *Economica, 80*, 118–130.

Fleurbaey, M., Peragine, V., & Ramos, X. (2017). Ex post inequality of opportunity comparisons. *Social Choice and Welfare, 49*(3), 577–603.

Kwasi Fosu, A. (2018). Economic structure, growth, and evolution of inequality and poverty in Africa: An overview. *Journal of African Economies, 27*(1), 1–9.

Gamboa, L. F., & Waltenberg, F. D. (2012). Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006–2009. *Economics of Education Review, 31*(5), 694–708.

Gamboa, L. F., & Waltenberg, F. D. (2015). Measuring inequality of opportunity in education by combining information on coverage and achievement in PISA. *Educational Assessment, 20*(4), 320–337.

Golley, J., & Kong, S. T. (2018). Inequality of opportunity in China's educational outcomes. *China Economic Review, 51*, 116–128.

Jaoul-Grammare, M., & Magdalou, B. (2017). Opportunities in higher education: An application to France. *Annals of Economics and Statistics/Annales D'Économie Et De Statistique, 111–112*, 295–325.

Kobus M., Kapera M., & Peragine V. (2020). *Measuring multidimensional inequality of opportunity* (UNU WIDER Working Paper N. 2020/19).

Lasso De La Vega, C., Lekuona, A., & Orbe, S. (2020). Re-examining the inequality of opportunity in education in some European countries. *Applied Economics Letters, 22*(7), 544–548.

Lefranc, A., Pistolesi, N., & Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France. *Journal of Public Economics, 93*(11–12), 1189–1207.

Li Donni, P., Rodriguez, J. G., & Rosa Dias, P. (2015). Empirical definition of social types in the analysis of inequality of opportunity: A latent classes approach. *Social Choice and Welfare, 44*(3), 673–701.

Li Donni, P., Peragine, V., & Pignataro, G. (2014). Ex-ante and Ex-post measurement of equality of opportunity in health: A normative decomposition. *Health Economics, 23*(2), 182–198.

Marrero, G. A., & Rodríguez, J. G. (2011). Inequality of opportunity in the United States: Trends and decomposition. In *Research on economic inequality* (Vol. 19, chapter 9).

Marrero, G. A., & Rodriguez, J. G. (2012). Inequality of opportunity in Europe. *Review of Income and Wealth, 58*(4), 597–621.

Marrero, G. A., & Rodriguez, J. G. (2013). Inequality of opportunity and growth. *Journal of Development Economics, 104*(1), 107–122.

Milanovic, B. (2015). Global inequality of opportunity: How much of our income is determined by where we live? *Review of Economics and Statistics, 97*(2), 452–460.

Paes de Barros, R., Ferreira, F. H. G., Molinas Vega, J. R., & Saavedra Chanduvi, J. (2009). *Measuring inequality of opportunities in Latin America and the Caribbean*. The World Bank.

Palmisano, F., Peragine, V., & Biagi, F. (2022). Inequality of opportunity in tertiary education in Europe. *Research in Higher Education*, *63*, 514–565.

Peragine, V. (2002). Opportunity egalitarianism and income inequality: The rank dependent approach. *Mathematical Social Sciences, 44*(1), 45–64.

Peragine, V. (2004). Ranking income distributions according to equality of opportunity. *The Journal of Economic Inequality, 2*(1), 11–30.

Peragine, V., & Serlenga, L. (2008). Higher education and equality of opportunity in Italy. In *Research on economic inequality*. Emerald Group Publishing Limited.

Pérez-Mesa, D. Marrero, G. A., & Darias-Curvo, S. (2020). *Child health inequality and opportunities in Sub-Saharan Africa* (ECINEQ WP Series N. 2020–557).

Pistolesi, N. (2009). Inequality of opportunity in the land of opportunities, 1968–2001. *Journal of Economic Inequality, 7*, 411–433.

Piraino, P. (2015). Intergenerational earnings mobility and equality of opportunity in South Africa. *World Development, 67*, 396–405.

Ramos, X., & Van de Gaer, D. (2020). Is inequality of opportunity robust to the measurement approach? *Review of Income and Wealth, 67*(1), 18–36.

Ramos, X., & Van de gaer, D. (2016). Empirical approaches to inequality of opportunity: Principles, Measures, and evidence. *Journal of Economic Surveys, 30*(5), 855–883.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Roemer, J. (1993). A pragmatic theory of responsibility for the Egalitarian Planner. *Philosophy & Public Affairs, 22*(2), 146–166.

Roemer, J. (1998). *Equality of opportunity*. Harvard University Press.

Roemer, J. E., & Trannoy, A. (2016). Equality of opportunity. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of income distribution* (Vol. 2A, pp. 217–300). Elsevier.

Salehi-Isfahani, D., Hassine, N. B., & Assaad, R. (2014). Equality of opportunity in educational achievement in the Middle East and North Africa. *The Journal of Economic Inequality, 12*(4), 489–515.

Schütz, G., Ursprung, H. W., & Wößmann, L. (2008). Education policy and equality of opportunity. *Kyklos, 61*(2), 279–308.

Shimeles, A., & Nabassaga, T. (2018). Why is inequality high in Africa? *Journal of African Economies, 27*(1), 108–126.

Singh, A. (2012). Inequality of opportunity in earnings and consumption expenditure: The case of Indian men. *Review of Income and Wealth, 58*, 79–106.

Trannoy, A., Tubeuf, S., Jusot, F., & Devaux, M. (2010). Inequality of opportunities in health in France: A first pass. *Health Economics, 19*(8), 921–938.

Van de gaer, D. (1993). *Equality of opportunity and investment in human capital*. PhD Dissertation, Katholieke Universiteit Leuven.

Van de gaer, D., Vandenbossche, J., & Figueroa, J. L. (2014). Children's health opportunities and project evaluation: Mexico's oportunidades program. *The World Bank Economic Review, 28*(2), 282–310.

World Bank. (2006). *World development report 2006: Equity and development*. World Bank.

PART III

Index Numbers and International Comparisons of Prices and Real Expenditures

CHAPTER 9

Framing Measurement Beyond GDP

Paul Schreyer

Introduction

Sound economic, social and environmental measurement relies on sound conceptual frameworks that provide relevant evidence. New social, environmental and economic issues have surfaced over the last two decades or so. More recently, policy makers' desire to 'Build Back Better' from the pandemic has put additional weight on the need for measurement approaches *Beyond GDP* or rather *GDP and Beyond*.[1]

[1] For a general discussion about GDP, its interpretation and uses, see, for instance, Stiglitz et al. (2009), Schreyer (2016), Hoekstra (2019), Heys et al. (2019), and Deaton and Schreyer (2021).

This article draws on a document provided as input to the 2020 meeting of the OECD Council at Ministerial Level (see OECD 2020). Opinions expressed are those of the author and do not necessarily reflect the views of the OECD or its Members.

P. Schreyer (✉)
OECD, Paris, France
e-mail: Paul.SCHREYER@oecd.org

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_9

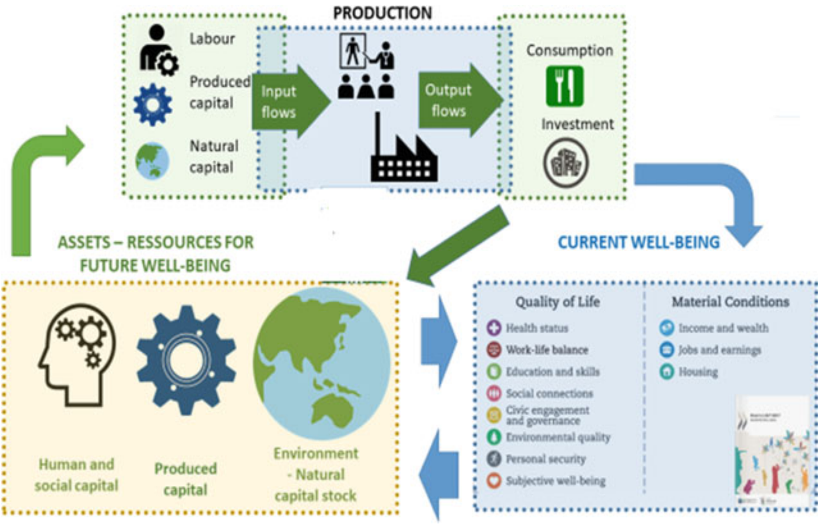

Fig. 9.1 Three spheres (*Source* OECD [2020])

At the risk of over-simplification, these issues can be addressed along three distinct but connected areas or spheres, each with its specific questions and corresponding measurement response (Fig. 9.1). They are the production (and market) sphere, the (*current*) well-being sphere and the asset (or sustainability) sphere that provides the resources for *future* production and well-being. There is significant value in measuring each of these spheres and the interactions between them. The three-tier framework is not new—its basic features can be found in Stiglitz et al. (2009), in the OECD's approach towards measuring Green Growth and Well-Being, in a number of national statistical publications and at least in its very basic structure, in a large body of academic literature around the theory of environment-economic measurement.[2]

The main purpose of what follows is sketching a measurement framework and drawing conclusions for the interpretation of well-known measures such as GDP or productivity growth. We also put forward some

[2] Basic features can be found in Weitzman's (1976) welfare interpretation of Net Domestic Product. Dasgupta (2009, 2021) are other examples. Fleurbaey and Blanchet (2013) provide an in-depth theoretical treatment.

measurement proposals that are feasible and relevant. Not everything that can be captured conceptually can be measured in practice—at reasonable cost—and not everything that can be aggregated in concept can be aggregated in practice. An example of the former is measuring societal shadow prices for all types of relevant assets, an example for the latter is coming up with a single measure that aggregates across all relevant dimensions of current well-being. A pragmatic approach is needed. The measurement agenda *can* be advanced and should be advanced to capture key aspects of the economy, society and the environment.

Production Sphere: What Gets in and What Comes Out of the "Factory Gates"

To characterise the production sphere, we consider a single-valued (for simplicity) non-negative aggregate measure of output, Q, volume GDP. And we consider the following (primary) inputs: L, to refer to labour input, measured as a single aggregate of hours worked; $S_K \equiv [S_{K1}, \ldots, S_{K,m-1}]$ to refer to $m - 1$ types of capital services flows from (mostly produced) capital stocks $K \equiv [K_1, \ldots, K_{m-1}]$ within the asset boundary of the System of National Accounts (see European Commission et al., 2009). These comprise produced assets (such as machinery, equipment or intellectual property assets) and some non-produced assets, in particular land and subsoil assets (such as minerals or petrol in the ground). All of the above inputs are market inputs, with service flows acquired through market transactions or through ownership over assets.

Non-market ecosystem services For the purpose at hand we extend the scope of factors that explicitly enter the production sphere and specify a flow of those ecosystem services $S_N \equiv [S_{N1}, \ldots S_{Nn}]$ that also shape the production sphere, but that are not subject to market transactions. The SEEA (2014) defines ecosystem services as " […]the multitude of resources and processes that are generated by ecosystem assets: collectively, these flows to people are referred to as ecosystem services. […] Flows of ecosystem services may relate either to flows of natural inputs from the environment to the economy (e.g., from the logging of timber resources) or to flows of residuals to the environment (e.g., emissions and waste) due to economic and other human activity. Flows of both natural inputs and residuals can impact on ecosystem assets, including on their structure, composition, processes, functions and biodiversity"

(SEEA, 2014, paragraph 2.14, p. 16). Some of these ecosystem services are thus the subject of market transactions (e.g. flows of timber resources) and would be captured as part of S_K whereas others would not be part of the standard national accounting system, such as emissions. Ecosystem services are rarely referred to in the measurement of production and productivity but correspond to what is generally meant in debates about 'the environment'. They will play a role for our conclusions on the measurement agenda and interpretation of sources of growth. Just like service flows from fixed assets (such as machinery) are related to a stock of assets, ecosystem services constitute flows from ecosystem assets such as forests, lakes or deep-sea floors.[3]

Ecosystem assets are defined spatially[4] and so are the associated ecosystem services, comprising provisioning services, regulating and maintenance services as well as cultural services. These go well beyond what enters directly as an input into producing units within the boundaries of the SNA. For instance, water filtration that helps crop yields provides direct input into the production of Q. Other ecosystem services such as cultural services (visual amenities or religious functions associated with nature) do not constitute input into production unless they are an enabling factor, for example, for the tourism industry. Provisioning services of timber from forests may fall under either S_N or S_K, depending whether economic ownership is exerted over the forest (in which case it is part of the SNA asset boundary) or not. Any particular type of ecosystem asset will provide a bundle of services to firms, households or government. But there is no unique mapping between flows of ecosystem services and particular ecosystem assets. As can already be gathered here, ecosystem services are complex, dynamic, spatially defined and potentially cross-border.[5] We shall return to these complexities later on and take it for the moment that production possibilities can be represented by a technology

[3] See SEEA (2021) for a full description.

[4] "Each ecosystem asset has a range of ecosystem characteristics - such as land cover, biodiversity, soil type, altitude and slope, climate - which describe the operation and location of the ecosystem. Some of these characteristics may be considered relatively fixed (e.g. slope and altitude), while others may be more variable (e.g. rainfall, land cover and biodiversity)" (SEEA, 2014, paragraph 2.12, p. 16).

[5] Globalised production in international value chains, coupled with digitalisation often entail cross-border flows of capital services from intangible assets such as intellectual property products. Several of the characteristics that apply to ecosystem services S_N are then equally valid for the measurement of inputs into subsidiaries of multi-national companies:

set ϕ^t such that:

$$\phi^t \equiv \left[(Q, L, S_K, S_N) : (L, S_K, S_N) \text{ can produce } Q \text{ in period } t\right]. \quad (9.1)$$

.

Measured productivity is conditional For available flows of ecosystem services, for given wage rates w and user costs of economic assets, u_K, producers combine L and S_K in a cost-minimising way to produce Q. To save on notation, we let $X \equiv [L, S_{K1}, ...S_{Km-1}]$ denote the m-valued vector of combined labour L and capital S_K inputs (other than ecosystem services), along with the m-valued vector of input prices $p_X \equiv [w, u_{K1}, ..., u_{Km-1}]$. Next, consider a conditional cost function[6] γ^t, defined as:

$$\gamma^t(Q, p_X, S_N) \equiv \min_X \left[p_X \cdot X \mid (Q, X, S_N) \in \phi^t\right] = p_X \cdot X. \quad (9.2)$$

γ^t thus reflects the minimum cost of producing Q, given a vector of input prices, and conditional on a level of entirely exogenous ecosystem services as well as autonomous technology available in period t. Minimum costs correspond to actual costs under the assumption of cost minimisation. We use $p_X \cdot X$ to denote the inner product of prices and quantities: $p_X \cdot X \equiv \sum p_{Xi} X_i$.

To keep things simple (but without significant consequences for the points to follow) we assume constant returns to scale and homotheticity in ecosystem services so that γ^t can be written as $\gamma^t = Q^t \mu^t(p_X)/\xi(S_N)$ where $\mu^t(p_X)/\xi(S_N)$ are unit costs of producing Q, non-decreasing in p_X and non-increasing in S_N. Thus, rising input prices increase unit costs, and more ecosystem services reduce them. Productivity growth between two periods $t = 0, 1$ can now be expressed as the change in cost for given input prices and environmental variables. A family of (inverted)

for example, the free use of a design or patent by a subsidiary has the form of a non-market transaction. In this context, Blanchet (2020) has argued that there is effectively no analytical way of capturing the contribution of individual factors of production from a domestic perspective and attention should be turned to measures of income, rather than GDP.

[6] This resembles a restricted cost function, as established by Lau (1976) and McFadden (1978). However, restricted cost functions were put forward to address situations where some (market) inputs are fixed or can only be adjusted in the longer term. This does not apply in the case of ecosystem services or free intellectual property assets.

productivity indices is given by $\Pi(Q, p_X, S_N) \equiv \frac{\gamma^0(Q, p_X, S_N)}{\gamma^1(Q, p_X, S_N)}$. Given the simplifying assumptions above,

$$\Pi(Q, p_X, S_N) \equiv \frac{\gamma^0(Q, p_X, S_N)}{\gamma^1(Q, p_X, S_N)} = \frac{Q\mu^0(p_X)/\xi(S_N)}{Q\mu^1(p_X)/\xi(S_N)} = \frac{\mu^0(p_X)}{\mu^1(p_X)}. \quad (9.3)$$

Two natural choices to evaluate the productivity index in (9.3) are with prices p_X^1 and p_X^0. We choose a geometric average to obtain:

$$\Pi(p_X^1, p_X^0) = \left[\frac{\mu^0(p_X^1)}{\mu^1(p_X^1)} \frac{\mu^0(p_X^0)}{\mu^1(p_X^0)}\right]^{1/2} = \frac{\mu^0(p_X^0)}{\mu^1(p_X^1)} \left[\frac{\mu^1(p_X^1)}{\mu^1(p_X^0)} \frac{\mu^0(p_X^1)}{\mu^0(p_X^0)}\right]^{1/2}. \quad (9.4)$$

Assume that the unit cost function $\mu^t(p_X)$ has a translog form (introduced by Christensen et al. (1971) and generalised by Diewert (1974). Diewert (1976) has shown that this flexible functional form approximates an arbitrary cost function to the second degree and the input price index on the right-hand side of (9.4) can be exactly represented by a Törnqvist index of the form $P_X^T \equiv \prod_{i=1}^m (p_{Xi}^1/p_{Xi}^0)^{0.5(v_i^1+v_i^0)}$ where $v_i^t \equiv \frac{p_{Xi}^t X_i^t}{p_X^t \cdot X^t}$ for $t = 0, 1$ is the cost share of each market input.

Next, expand (9.4) to arrive at a standard form of productivity measurement:

$$\begin{aligned}\Pi(p_X^1, p_X^0) &= \frac{\mu^0(p_X^0)}{\mu^1(p_X^1)} P_X^T(p_X^1, p_X^0)\\ &= \frac{Q^0\mu^0(p_X^0)/\xi(S_N^0)}{Q^1\mu^1(p_X^1)/\xi(S_N^1)} P_X^T(p_X^1, p_X^0) \frac{Q^1/\xi(S_N^1)}{Q^0/\xi(S_N^0)}\\ &= \frac{p_X^0 \cdot X^0}{p_X^1 \cdot X^1} P_X^T(p_X^1, p_X^0) \frac{Q^1/\xi(S_N^1)}{Q^0/\xi(S_N^0)}\end{aligned}$$

$$\Pi(p_X^1, p_X^0)\frac{\xi(S_N^1)}{\xi(S_N^0)} = \frac{Q^1}{Q^0} \Big/ \left[\frac{p_X^1 \cdot X^1}{p_X^0 \cdot X^0} \Big/ P_X^T(p_X^1, p_X^0)\right]. \quad (9.5)$$

The right-hand side of (9.5) looks remarkably standard—productivity growth is the ratio between real output growth and 'real' (deflated) change in input values. But the left-hand side shows the dependence of this growth accounting equation on the ecosystem variables via $\xi(S_N^1)/\xi(S_N^0)$. The point here is that even with the rather strong assumption of homotheticity of costs in the environmental variables (which implies dependence of the input price index $P_X^T(p_X^1, p_X^0)$ and the theoretical productivity variable $\Pi(p_X^1, p_X^0)$ on input prices only), the standard way of computing multi-factor productivity—call it $MFP \equiv \frac{Q^1}{Q^0}/\left[\frac{p_X^1 \cdot X^1}{p_X^0 \cdot X^0}/P_X^T(p_X^1, p_X^0)\right] = \Pi(p_X^1, p_X^0)\frac{\xi(S_N^1)}{\xi(S_N^0)}$—implies a conditionality on environmental variables.[7] Thus, if the production sphere benefits from a rising flow of ecosystem services ($\frac{\xi(S_N^1)}{\xi(S_N^0)} \geq 1$), measured productivity growth *MFP* will be overstated: $MFP(p_X^1, p_X^0, S_N^1, S_N^0) \geq \Pi(p_X^1, p_X^0)$. Conversely, if production becomes less intensive in its use of ecosystem services, measured productivity growth is understated. Also, the production process might—as a by-product—enhance ecosystem assets, for instance when agricultural activity maintains or improves the landscape or enhances soil fertility through organic production techniques. Then S_N would take a negative sign and constitute a (non-remunerated) output. Whether measured productivity growth is over- or understated depends again on whether the flow of such by-products increases or decreases.

The link with ecosystem services here has been the simplest conceivable in the sense that their usage is free and there are no binding constraints or regulations. A richer framework would introduce constraints on the use of ecosystem services and recognise the fact that reducing the use of these services requires resources and implies foregoing market output. Pittman (1983), Färe et al. (1993), Vanoli (1995), Brandt et al. (2014) and Cardenas Rodriguez et al. (2018) are examples of such work. For instance, the latter value the shadow prices of air emissions across a set of OECD countries. Such explicit adjustments or rather decompositions of a potential GDP or MFP into observed (good) final outputs and the outputs devoted to abatement are possible and useful. Note, however, that by their very nature they embrace a *producer* viewpoint so that ecosystem services are valued from a private, not a social perspective.

To summarise, there is no suggestion here that price and productivity measures should systematically be adjusted for the entire set of ecosystem services. Indeed, the difficulties of measuring and valuing

[7] If conditions are relaxed, e.g. to non-homotheticity of ecosystem services in costs, the input price index the shift in unit costs would directly depend on S_N, reinforcing the point made here.

ecosystem services will settle the question. But explicit recognition of *selected* ecosystem services is possible and useful and can be pursued at reasonable statistical cost. And if no such adjustment is undertaken, it is important to remember that our standard productivity measurement tools will inevitably pick up such effects, and, in the light of their rising importance, pull our *MFP* metrics further away from a traditional 'engineering' technology interpretation.

Non-market production of households We have so far glossed over an important element in production, non-market activities carried out by households. With the exception of owner-occupied housing (where the System of National Accounts foresees an imputation for the value of housing services that house owners provide to themselves), other non-market services produced by households are not included in GDP. This includes, for instance, teaching services provided by parents to their children, nursing services provided to infirm relatives or friends, cooking or gardening on ones' own premises all of which are acts of production and yet outside GDP by convention.

There is a long tradition of estimating the value of the non-market production of households, starting in the 1930s (Reid, 1934). A basic requirement is the availability of time use surveys, not necessary a matter of course. Valuation of hours of labour input at home entails other complications, with the replacement cost and the opportunity cost approaches as standard methods (for a recent application to OECD countries see for instance Van de Ven et al., 2018). However, Schreyer and Diewert (2014) have shown that the choice for valuing different types of household production depends on the socio-economic characteristics of the household—for example, whether or not it is constrained in its supply of labour on the market. Thus, the valuation of unpaid household work is not a settled matter, and numbers are large. Regular evaluations in a standardised accounting framework without, however, an inclusion in GDP would seem the right way forward to recognise this important aspect of the production sphere. Indeed, as these activities clearly affect people's well-being outcome (health, education, social connections, etc.), they also connect naturally to the well-being sphere.

Conclusion 1 GDP, production and productivity remain useful measures for governments' fiscal and monetary policies and macro-economic monitoring. But they need careful interpretation as ecosystem services increasingly interact with the production sphere and many of our standard measures are conditioned on these services. GDP is also oblivious to most production activity by private households outside the marketplace, it takes no account of how incomes are distributed and it is sometimes driven by income from intangible assets such as intellectual property that are moved between jurisdictions to minimise corporate tax burdens (Deaton & Schreyer, 2021).

Accounting system (9.1) and (9.2) constitute a simplified accounting system, with the production side reflected in (9.1) and the income side captured by (9.2). The national accounting system is complete by adding in the expenditure side where the current value of GDP ($Y^t \equiv P_Q^t Q^t$ with P_Q^t as the GDP deflator) equals private and government aggregate volume consumption C^t (with a corresponding deflator P_C^t) and a vector of volume investments $I^t \equiv [I_1^t, \ldots I_{m-1}^t]$. We ignore exports and imports and taxes for simplicity of exposition. Gross investment in market assets I^t corresponds to the change in the capital stock $\Delta K = K^t - K^{t-1}$ plus depreciation (or depletion) $D(K^{t-1})$, all valued at prices $p_I^t \equiv [p_{I,1}^t, \ldots p_{I,m-1}^t]$.

We pause here to underline that while measures of depreciation or consumption of fixed capital (the loss of value of *produced* assets as they are employed in production) are generally available, this is much less the case for depletion (or discoveries) of *non-produced* assets even if these are inside the SNA asset boundary, in particular subsoil assets. A first and important task is measurement of such depletion and discoveries and the SEEA (2014, 2021) provides all the necessary guidance. Ecosystem services reduce or enhance ecosystem assets and so bear a resemblance to the notions of depreciation, depletion or discoveries. However, as pointed out above, they are complex to gauge and typically there are no meaningful market valuations to go by. We, therefore, just depict the stock-flow relationship in physical units and formulate it as a general, possibly non-linear, many-to-many mapping $f^t(\cdot)$ that also allows for other factors Ω to affect ecosystem assets.

$$Y^t \equiv P_Q^t Q^t = P_C^t C^t + p_I^t \cdot I^t = p_X^t \cdot X^t = w^t L^t + u_K^t \cdot S_K^t$$

$$\text{where } p_I^t \cdot \Delta K = p_I^t \cdot I^t - p_I^t \cdot D(K^{t-1}), \text{ and}$$
$$\Delta N = f^t(S_N^t, N^{t-1}, \Omega^t). \tag{9.6}$$

Equation (9.6) constitutes the link between the production sphere and the asset sphere, further explored in Sect. "Asset Sphere: The Resources for Future Well-Being". Equation (9.6) also constitutes a link to the well-being sphere in Sect. "Well-being Sphere: What Shapes People's Lives?" because it gives immediate rise to net national income (NI), the flow of income adjusted for depreciation (and net income transfers from abroad—ignored here for simplicity). Relevance for economic well-being arises because, in a very simple setting, NI captures aspects of both current and future consumption in particular when expressed in real terms after deflation with a consumption price index:

$$NI^t \equiv Y^t - p_I^t \cdot D(K^{t-1}) = P_C^t C^t + p_I^t \cdot \Delta K$$
$$NI^t / P_C^t = C^t + \frac{p_I^t}{P_C^t} \cdot \Delta K \cdot \tag{9.7}$$

We also see that when capital is exactly kept intact ($\Delta K = 0$), the maximum possible consumption equals real net income which corresponds to the basic notion of Hicksian Income (Hicks, 1940.[8]). Weitzman (1976) demonstrated how in a simple closed economy without technical progress real net income is proportional to the present discounted value of consumption that the economy is able to produce, thus also giving NI^t / P_C^t meaning as a dynamic measure of economic well-being—a proposal discussed in many places, starting with Usher (1976). Sefton and Weale (2006) demonstrated that real savings (which equals real net investment in a closed economy), expressed in consumption equivalents, is the proper indicator of changes to inter-temporal well-being. All that said, the assumptions needed to confidently interpret real net income as a true measure of economic well-being are strong, and include reliance on inter-temporal general equilibria, leaving aside aspects of substitutability between consumption and investment and the evolution of future productivity trends. Also, aggregate measures of income are oblivious to its distribution among individuals and households which

[8] "However, if we do decide to include saving in our Welfare index, the appropriate concept of individual income can be nothing else but what the individual *thinks* he can consume without making himself worse off" (p. 123).

will form a central theme as we turn to the well-being sphere in the next section.

With a more pragmatic and less ambitious interpretation as maximum aggregate current consumption possibilities once allowance is made for replacement investment and depletion—but ignoring other non-market assets and therefore questions of sustainability—real net income remains a very useful concept, especially if:

- the set of SNA capital measures K is as complete as possible and includes in particular non-produced, non-financial assets such as subsoil assets and land. One notes that while the quantity of land is more or less fixed, its quality is not. Land degradation and land improvements are thus part of depreciation and capital formation, respectively;
- deflation of NI is achieved with a consumption price index;
- national rather than domestic income forms the basis to correct for (actual and imputed) international transfers of income.

Conclusion 2 Real net national income is a measure that constitutes a first step towards capturing average current economic well-being. While available for most countries, its empirical basis needs improving through updated measures of depreciation, and full consideration of depletion, discoveries and quality change of those non-produced assets that are already part of the national accounts asset boundary, in particular subsoil assets and land.

Well-being Sphere: What Shapes People's Lives?

The notion of well-being has gained increasing traction over the last twenty years as an agenda for research, measurement and policy. Particular early impetus had come through the Human Development Index (UNDP, 2020) and through the recommendations of the Stiglitz-Sen-Fitoussi Commission (Stiglitz et al., 2009) but the body of related research is large (see Jorgenson, 2018 for an overview and Landefeld et al., 2020 for forthcoming work in the United States Bureau of Economic Analysis). The OECD's (2015) empirical work with its

How's Life indicator dashboard has also been on the forefront. It defines current well-being in terms of *Material Living Conditions* and *Quality of Life*, captured through eleven dimensions that shape people's lives. These dimensions are income and wealth, jobs and earnings, housing, health, work-life balance, skills, social connections, civic engagement and governance, environmental quality, personal security and subjective well-being. A key feature of well-being measures is also that they go beyond averages and consider the distribution of outcomes across individuals and households. Inequalities are central to measuring well-being, both of the material and of the quality of life sort. We thus see that the aggregate income and (economic) wealth dimensions that link back to the production sphere are but a small part of the determinants of current well-being.

Distribution of well-being To capture current well-being more formally, define a utility function for a household (or type of household) h such that $U_h = U_h(C_h, S_{Nh}, Z_h)$[9] where C_h is household's h current consumption of market products, $S_{Nh} \equiv [S_{Nh,1}, ...S_{Nh,n}]$ is its consumption of non-market ecosystem services and $Z_h \equiv [Z_{h,1}, ...Z_{h,l}]$ depicts the vector of other *Quality of Life* outcomes. Household h's minimum expenditure conditional on S_{Nh} and Z_h and a vector of consumer prices $p_C \equiv [p_{C,1}, ...p_{C,c}]$is then given by:

$$E_h(U_h, p_C, S_{Nh}, Z_h) \equiv \min_{C_h}[p_C \cdot C_h \mid U_h(C_h, S_{Nh}, Z_h) \geq U_h] = p_C \cdot C_h. \tag{9.8}$$

An important step in the task of measuring economic well-being is breaking down income, expenditure and consumption aggregates (9.6 and 9.7) by category of household, thus relaxing the assumption of a representative consumer. While there is a long tradition of measuring income and consumption by individual or by household through surveys (and tax records), these statistics are not normally sufficient to achieve a breakdown that is consistent with the national accounts and a series of adjustments are required (Jorgenson & Schreyer, 2017). Recent research has already come forward with interesting results, including Zwijnenburg

[9] It is assumed that U_h is continuous and increasing in the components of C_h, S_{Nh} and Z_h, and is concave in the components of C_h.

et al. (2017), Piketty et al. (2018) and U.S. Bureau of Economic Analysis (2020) and several national statistical offices. Figure 9.2 shows an example.

Measured living standards are conditional There is no space here to discuss the many important practical questions that need to be resolved to achieve national accounts consistency and we refer to the above publications. But there is a conceptual point to be made here, namely that measures of the distribution of income or consumption, even when fully consistent with national accounts concepts remain approximations to living standards, including material living standards, because they are conditional on 'environmental variables' S_{Nh}, Z_h. This is readily seen by comparing consumption expenditures of two (groups of) households h and h' (e.g. the first and fifth quintile in the income distribution or households in two regions), using (9.8):

$$\frac{p_C \cdot C_h}{p_C \cdot C_{h'}} = \frac{E_h(U_h, p_C, S_{Nh}, Z_h)}{E_{h'}(U_{h'}, p_C, S_{Nh'}, Z_{h'})} \tag{9.9}$$

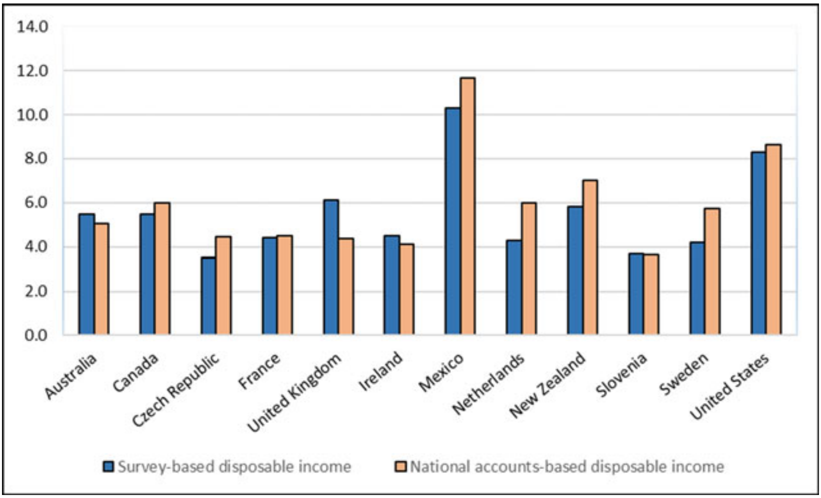

Fig. 9.2 Disposable income fifth over first quintile: survey-based and national accounts-based (*Note* The graph shows disposable income per consumption unit for the fifth quintile relative to the adjusted disposable income for the first quintile. *Source* Zwijnenburg et al. [2021])

As can be seen from (9.9), consumption expenditures of the two households reflect both the levels of utility $U_h, U_{h'}$—which is what we would like to capture—*and* the household-specific environmental variables. We have assumed here for simplicity that all consumers h and h' face the same prices p_C but household-specific prices could be accommodated.[10]

A true comparison of living standards based on consumption expenditure, call it $Q_{Uhh'}$, needs to be contingent on reference variables S_N, Z that are identical for the households under comparison. Thus, $E_h(U_h(C_h, S_{Nh}, Z_h), p_C, S_N, Z)$ is the amount which, if made available to the consumer when facing prices p_C and the reference quantities S_N, Z, would make the consumer just as well off as at $U_h(C_h, S_{Nh}, Z_h)$. This is of course nothing but Samuelson's (1974) and Samuelson and Swamy's (1974) Money Metric Utility, extended to include environmental variables as in Willig (1981), Blundell et al. (1994), and Fleurbaey and Gaullier (2009). A living standard comparison $Q_{Uhh'}$ between two households is then given by

$$\begin{aligned} Q_{Uhh'} &\equiv \frac{E_h(U_h, p_C, S_N, Z)}{E_{h'}(U_{h'}, p_C, S_N, Z)} \\ &= \frac{E_h(U_h, p_C, S_{Nh}, Z_h)}{E_{h'}(U_{h'}, p_C, S_{Nh'}, Z_{h'})} \\ &\quad \left[\frac{E_h(U_h, p_C, S_N, Z)}{E_h(U_h, p_C, S_{Nh}, Z_h)} \frac{E_{h'}(U_{h'}, p_C, S_{Nh'}, Z_{h'})}{E_{h'}(U_{h'}, p_C, S_N, Z)}\right] \\ &= \frac{p_C \cdot C_h}{p_C \cdot C_{h'}} \beta \\ \text{with } \beta &\equiv \left[\frac{E_h(U_h, p_C, S_N, Z)}{E_h(U_h, p_C, S_{Nh}, Z_h)} \frac{E_{h'}(U_{h'}, p_C, S_{Nh'}, Z_{h'})}{E_{h'}(U_{h'}, p_C, S_N, Z)}\right] \end{aligned} \tag{9.10}$$

For well-being comparisons, consumption expenditure (or income) comparisons between two households thus need to be adjusted by a factor β which corrects for the differences of each household's situation compared to reference conditions S_N and Z. More precisely β is an index of the expenditure that household h would have to be compensated for, given its distance to references conditions compared to the compensation

[10] This is in particular needed for international comparisons because price levels differ between countries. Rao (2016) provides an overview of comparisons of economic well-being in conjunction with purchasing power parities.

of household h'. Each ratio in β constitutes an index of money metric utility or willingness to pay, reflective of household specific preferences and the household-specific situation with regard to the reference variables. We note that for β to equal unity so that the simple ratio of consumption expenditures is an accurate measure of relative well-being of the two households, preferences need to be identical and homothetic in the level of utility and the 'environmental' variables *and* each household needs to enjoy (or be subjected to) the same level of environmental variables.

Is it possible to measure expressions such as $E_h(U_h, p_C, S_N, Z)$? The answer is yes, although modelling is required and in general, the number of environmental variables that can be controlled for is limited. Examples include Fleurbaey and Gaulier (2009), Jones and Klenow (2016) and Boarini et al. (2022). The latter estimate measures of equivalent income for health (life expectancy) and jobs for a group of countries (see Fig. 9.3), allowing for heterogeneous preferences. In this case, the value of market consumption (or income) in a particular country is reduced by a monetary value that corresponds to the distance of a country's life expectancy from the world leader (Japan) times the (country-specific, revealed) willingness to pay for an extra year of life expectancy on average. Similarly, an equivalent income measure is constructed for the value of full employment (above and beyond the remuneration for work which is already captured by market income/consumption).

Conclusion 3 Constructing national accounts compatible measures of the distribution of consumption, income (and possibly wealth) by household is a key task ahead and a necessary input for a social welfare measure. Akin to GDP and productivity, such a welfare measure remains, however, conditional on the distribution of the outcome of other well-being dimensions across households. Integration of several—but likely not all—such dimensions into a broader, single measure of income or consumption is possible and worth pursuing when the theoretical basis for such a composite measure is solid. Other than that the conditionality of income and consumption comparisons on environmental variables needs to be kept in mind and it is often helpful to show these in separate dashboards.

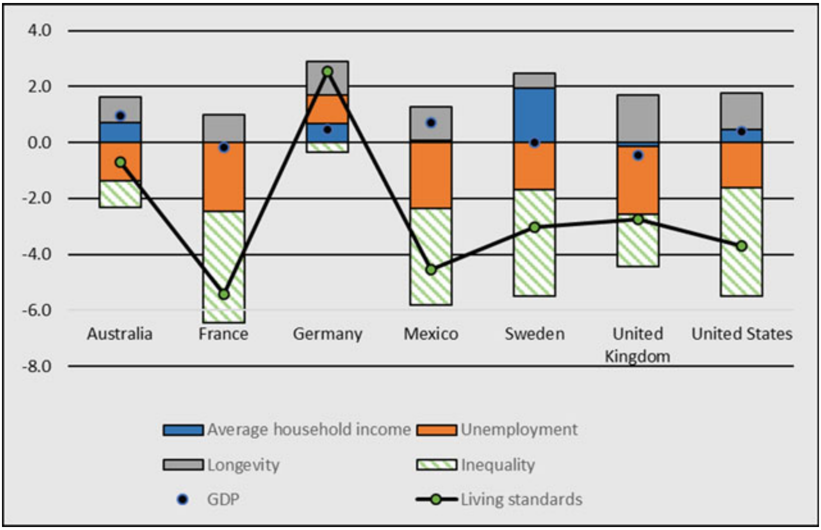

Fig. 9.3 Multi-dimensional living standards for bottom quintile of households Average annual percentage change 2008–2013 (*Note* The graph shows the contributions of average household disposable income, unemployment and longevity to living standards and an adjustment for aversion to inequality. Elements are weighted with their shadow prices, i.e. the willingness to pay to avoid unemployment [capturing the value of jobs above and beyond the income generated by them] and the willingness to pay to reduce mortality risk to the best performing country]. *Source* [Boarini et al., 2022]).

Asset Sphere: The Resources for Future Well-Being

The notion of economic, environmental or social sustainability requires invoking assets in a broad sense. While there are many specific definitions of sustainability, they all take an inter-temporal view and preserving wealth is a natural way of thinking about this. The 1987 Brundtland Report's definition of sustainable development as "[...] development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations, 1987). The OECD's work on well-being (OECD, 2015) understands sustainability as acting in a way that assets are preserved for future generations' well-being. The World Bank's *Changing Wealth of Nations* (World Bank,

2018) is a concrete effort towards valuing the level and development of countries' assets. And the 2015 UN Sustainable Development Goals and some of the associated indicators can also be understood as an effort to measure humanity's capacity to preserve economic, natural and social assets for future generations alongside measures of current well-being and its distribution.

Accounting prices—theoretically strong but very hard to implement Indeed, the change in the value of comprehensive wealth is one way of defining and conceptualising sustainability, and a body of theoretical papers has explored this idea (for an overview see Dasgupta, 2009). At the heart of the matter is an extension to all types of assets—market and non-market—of the ideas around real savings or net investment as a measure of changes in inter-temporal economic well-being (see Sect. 'Production Sphere: What Gets in and What Comes Out of the "Factory Gates"'). An inter-temporal social welfare function that augments the environment-economy accounting system as presented in (9.6) is formulated as the discounted value of future utilities of each of the $h = 1, \ldots H$ households:

$$V^0 = \sum_{t=0}^{\infty} W(U_1(C_1^t, S_{N1}^t), \ldots U_H(C_H^t, S_{NH}^t))(1+r)^{-t} \tag{9.11}$$

Taking account of the stock-flow relationships portrayed in (9.6), Dasgupta (2009) and Arrow et al. (2003) introduce the notion of resource allocation mechanisms, i.e. future paths of the economic-environmental system, and demonstrate that it is possible, by recursive reasoning, to map particular resource allocation mechanism onto today's capital stocks so that $V^0 = V^0(C_1^0, \ldots, C_H^0, K^0, N^0, \alpha)$ where α is a particular resource allocation mechanism.[11] No optimal behaviour is required to introduce this concept. 'Accounting prices' (i.e. marginal social valuations) of capital stocks are introduced, defined as $p_{AKi} \equiv \frac{\partial V^0(C_1^0, \ldots, C_H^0, K^0, N^0, \alpha)}{\partial K_i}$, $p_{ANi} \equiv \frac{\partial V^0(C_1^0, \ldots, C_H^0, K^0, N^0, \alpha)}{\partial N_i}$. The changes in stocks today (presented here in continuous time for simplicity), each valued at

[11] Note that we have ignored quality of life variables Z here, for simplicity. With some stretch of imagination, each quality of life dimension (such as health, personal safety) could be conceived as either an additional type of capital (such as human capital or social capital) or as a service associated with such capital and integrated into the theoretical concept.

accounting prices, then gauge the change in inter-temporal social welfare, depending on whether $dV^0 \lesseqgtr 0$ (Dasgupta, 2009):

$$dV^0 = \sum_{i=1}^{m-1} p_{AKi} dK^0 + \sum_{i=1}^{n} p_{ANi} dN^0. \tag{9.12}$$

Tracking changes in the asset base with accounting prices would thus seem to be the most important effort to pursue. But it also turns out to be the most challenging venture, in concept and in practice. There are at least two difficulties here.[12]

- Although 'only' present changes in assets need to be observed, their valuation with accounting prices requires projections of the future evolution of the socio-economic-environmental system, because the resource allocation mechanism α has to be described and evaluated. Conceptually, accounting prices reflect all the negative externalities associated with economic activities, missing markets and increase when stocks of capital approach 'tipping points'. Here we are in a different world from that normally inhabited by statistical offices—a world of scenario-building, horizon scanning and comprehensive modelling and forecasting. Resource requirements apart, this raises some important institutional issues.
- An indicator of sustainability needs to be based on comprehensive wealth, encompassing a broad set of assets, from produced machinery to human capital, social capital and natural assets. But what is the exact scope and how should it be measured? There are many borderline cases, and measurement issues abound, including:
  - Whether or not **health** ought to be recognised as a separate asset, and in addition to human capital is a matter of debate and makes a tremendous difference to results Arrow and Dasgupta (2003).
  - **Human capital** measurement methods are well established, in particular those in the tradition of Jorgenson and Fraumeni (1989) but one notes that they imply projections of future

[12] See Fleurbaey and Blanchet (2013) for a broader discussion of measurement and theoretical questions.

income. Another asset considered significant for the functioning of societies is **social capital**, i.e. the social norms, shared values and institutional arrangements that foster co-operation among population groups and the trust people have in others. For instance, OECD (2015)'s headline indicators on the level and evolution of social capital are based on surveys on the level of trust between individuals and in institutions. It would appear very difficult to develop accounting prices for social capital.

– Another key question is capturing and valuing **ecosystem assets**. Their deterioration or improvement represents a big part of what constitutes today's environmental concerns, and it has been pointed out earlier that ecosystem assets are complex, dynamic, they do not observe national boundaries and there is no simple mapping to ecosystem service flows. Modelling accounting prices that attach to the change in ecosystem assets would appear to be a tall order.

Pragmatism is the word Should we thus refrain from measuring and valuing assets and how they evolve over time? Definitely not. We do need relevant measures but pragmatism should reign and our ambitions to develop a robust single indicator of sustainability, or even non-sustainability need to be kept in check. The work by the World Bank (2018)—motivated by the idea that we need to look at a broad set of assets to get a sense of where sustainability is heading—is a good example of a pragmatic approach. The *Dasgupta Review* (2021) is a showcase for the many empirical and theoretical aspects that measurement of biodiversity and ecosystem services from biodiversity entail, even in physical terms. The OECD's *How's Life?* series (OECD, 2015) organises and presents available cross-country evidence on assets, to paint a picture where things are heading with produced, human, social and natural capital—see Fig. 9.4. No claim is made on comprehensiveness nor is there aggregation across assets.

	Headline indicator	OECD average and range, 2018 or latest available year	OECD average change since 2010	No. of countries consistently improving	No. of countries consistently deteriorating
Economic Capital	**Produced fixed assets** (USD per capita at 2010 PPPs)	POL ~36 500; OECD 31 ~119 000; NOR ~195 000	+11 percent points	23	3
	Financial net worth of general government (percentage of GDP)	GRC -143; OECD 36 -27; NOR 280	-4 percent. points	5	13
	Household debt (as a share of household net disposable income)	DNK 281; OECD 33 126; HUN 43	-3 percent. points	12	13
Natural Capital	**Greenhouse gas emissions** (CO^2 equivalent, domestic production, tonnes per capita)	AUS 22.5; OECD Total 11.9; COL 3.2	-1 tonne/ capita	22	2
	Material footprint (used raw material extracted to meet the economy's final demand, tonnes per capita)	LUX 102; OECD Total 25; MEX 9	+1.2 tonne/capita	3	16
	Red List Index of threatened species (0 = all species extinct; 1 = all species qualifying as least concern)	NZL 0.62; OECD 0.89; SWE 0.99	-0.01	13	2
Human Capital	**Educational attainment among young adults** (share of people aged 25-34 years with at least upper secondary education)	MEX 50; OECD 35 85; KOR 98	+2 percent. points	26	3
	Labour underutilization (share of unemployed, discouraged or underemployed workers in the labour force)	GRC 28; OECD 26 12; CZE 4	-4.8 percent. points	15	2
	Premature mortality (potential years of life lost due to medical conditions and fatal accidents per 100 000 inhabitants)	LVA ~8700; OECD 36 ~4600; CHE ~3000	-620 years lost	29	2
Social Capital	**Trust in others** (mean score on a scale from 0 – 10)	TUR 4.5; OECD 28 6.1; DNK 8.3	No time series		
	Trust in government (share of the population responding positively)	GRC 14; OECD 43; CHE 82	+3 percent. points	9	6
	Gender parity in politics (share of women in national parliament)	JPN 9; OECD 28; ISL 48	+2.6 percent. points	11	2

Fig. 9.4 Evolution of different types of capital in OECD countries (*Note* The snapshot depicts data for 2019, or the latest available year, for each indicator. The colour of the circle indicates the direction of change, relative to 2010, or the closest available year: consistent improvement is shown in blue, consistent deterioration in orange, no clear trend in grey and insufficient time series to determine trends in white. For each indicator, the OECD country with the lowest [on the left] and highest [on the right] well-being level are labelled, along with the OECD average [in black, and unless all 37 members are included detailing the number of countries in the average]. *Source* OECD [2011, 2013, 2015, 2017, 2020])

Conclusion 4 *The comprehensive social* valuation of economic, environmental and social assets is an excellent reference framework to reason about sustainability. But its empirical implementation with an ambition of providing a comprehensive, single indicator of sustainability raises more questions than it may answer. A pragmatic approach is called for: starting with a fuller implementation of SNA assets, improved physical measures of natural assets, spatially differentiated valuations of ecosystem services are good places to start. The latest version of the SEEA United Nations, European Union, Food and Agriculture Organization of the United Nations, Organisation for Economic Co-operation and Development, World Bank Group (2014) provides excellent guidance here. The measurement of human capital is also well-established and worth pursuing periodically.

References

Arrow, K. J., Dasgupta, P., & Mäler, K.-G. (2003). Evaluating projects and assessing sustainable development in imperfect economies. *Environmental and Resource Economics, 26*(4), 647–685.

Blanchet, D. (2020). What should the concept of domestic production mean in globalized economies? *Economie et Statistique/Economics and Statistics*, 517–518–519, 205–214. https://doi.org/10.24187/ecostat.2020.517t.2019

Blundell, R., Preston, I., & Walker, I. (1994). An introduction to applied welfare analysis. In R. Blundell, I. Preston, & I. Walker (Eds.), *The measurement of household welfare* (pp. 1–50). Cambridge University Press.

Boarini, R., Fleurbaey, M., Murtin, F., & Schreyer, P. (2022). Well-being during the great recession: New evidence from a measure of multi-dimensional living standards with heterogenous preferences. *Scandinavian Journal of Economics*, *124*(1), 104–138.

Brandt, N., Schreyer, P. , & Zipperer, V. (2014). *Productivity measurement with natural capital and bad outputs* (OECD Economics Department Working Papers, No. 1154). OECD Publishing. https://doi.org/10.1787/5jz0wh5t0ztd-en

Bureau of Economic Analysis. (2020). *Distribution of personal income*, Prototype statics. U.S. BEA. https://www.bea.gov/data/special-topics/distribution-of-personal-income

Cárdenas Rodríguez, M., Haščič, I., & Souchier, M. (2018). *Environmentally adjusted multifactor productivity: Methodology and empirical results for OECD and G20 countries* (OECD Green Growth Papers, No. 2018/02). OECD Publishing.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1971). Conjugate duality and the transcendental logarithmic production function. *Econometrica, 39*, 255–256.

Dasgupta, P. (2009). The welfare economic theory of green national accounts. *Environmental and Resource Economics, 42*(1), 3–38.

Dasgupta, P. (2021). *The economics of biodiversity: The Dasgupta review*. HM Treasury. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/962785/The_Economics_of_Biodiversity_The_Dasgupta_Review_Full_Report.pdf

Deaton, A., & Schreyer, P. (2021). GDP, wellbeing, and health: Thoughts on the 2017 round of the International Comparison Program. *Review of Income and Wealth*.

Diewert, W. E. (1974). Applications of duality theory. In M. D. Intrilligator & D. A. Kendrick (Eds.), *Frontiers of quantitative economics* (Vol. II, pp. 106–171). Amsterdam: North-Holland Publishing Co.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics, 4*, 115–145.

European Commission, IMF, OECD, World Bank, & United Nations. (2009). *2008 System of national accounts*. United Nations. http://unstats.un.org/unsd/nationalaccount/sna2008.asp.

Färe, R., Grosskopf, S., Knox Lovell, C. A., & Yaisawarng, S. (1993). Derivation of shadow prices for undesirable outputs: A distance function approach. *The Review of Economics and Statistics, 75*(2), 374–380.

Fleurbaey, M., & Blanchet, D. (2013). *Beyond GDP: Measuring welfare and assessing sustainability*. Oxford University Press.

Fleurbaey, M., & Gaulier, G. (2009). International comparisons of living standards by equivalent incomes. *Scandinavian Journal of Economics, 111*(3), 529–624.

Heys, R., Martin, J., & Mkandawire, W. (2019). *GDP and welfare: A spectrum of opportunity*. ESCoE DP 2019-16. https://www.escoe.ac.uk/publications/gdp-and-welfare-a-spectrum-of-opportunity/

Hicks, J. R. (1940). The valuation of the social income. *Economica, 7*(26) (May), 105–124.

Hoekstra, R. (2019). *Replacing GDP by 2030: Towards a common language for the well-being and sustainability community*. Cambridge University Press.

Jones, C. I., & Klenow, P. J. (2016). Beyond GDP: Welfare across countries and time. *American Economic Review, 106*(9), 2426–2457.

Jorgenson, D., & Fraumeni, B. M. (1989). The accumulation of human and nonhuman capital, 1948–84. In *The measurement of saving, investment, and wealth* (pp. 227–286). National Bureau of Economic Research. https://EconPapers.repec.org/RePEc:nbr:nberch:8121

Jorgenson, D. W. (2018). Production and welfare: Progress in economic measurement. *Journal of Economic Literature, 56*(3), 867–919.

Jorgenson, D. W., & Schreyer, P. (2017). Measuring individual economic well-being and social welfare within the framework of the system of national accounts. *The Review of Income and Wealth, 63*, S2, S460-S-477.

Landefeld, S., Villones, S., & Holdren, A. (2020). GDP and beyond: Priorities and plans. *Survey of Current Business, 100*(6), 1–34. https://apps.bea.gov/scb/2020/06-june/0620-beyond-gdp-landefeld.htm

Lau, L. J. (1976, February). A characterization of the normalized, restricted profit function. *Journal of Economic Theory, 12,* 131–63.

McFadden, D. (1978). Cost, revenue and profit functions. In M. Fuss & D. McFadden (Eds.), *Production economics: A dual approach to theory and applications* (Vol. 1, pp. 3–109). North-Holland.

OECD. *Framing the measurement of production, well-being and sustainability* SDD/CSSP(2020)2/REV1. https://one.oecd.org/document/SDD/CSSP(2020)2/REV1/en/pdf

OECD. (2011, 2013, 2015, 2017, 2020). *How's life?*. OECD Publishing. http://www.oecd.org/statistics/how-s-life-23089679.htm

Piketty, T., Saez, E., & Zucman, G. (2018). Distributional national accounts: Methods and estimates for the United States. *The Quarterly Journal of Economics, 133*(2) (May) 553–609.

Pittman, R. W. (1983). Multilateral productivity comparisons with undesirable outputs. *The Economic Journal, 93*(372), 883–891.

Rao, P. (2016). Welfare comparisons with heterogenous prices, consumption, and preferences. In M. Adler & M. Fleurbaey (Eds.), *The Oxford handbook of well-being and public policy*. Oxford University Press.

Reid, M. (1934). *The economics of household production*. Wiley.

Samuelson, P. A. (1974). Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. *Journal of Economic Literature, 12*, 1255–1289.

Samuelson, P. A., & Swamy, S. (1974). Invariant economic index numbers and canonical duality: Survey and synthesis. *American Economic Review, 64*, 566–593.

Schreyer, P. (2016). GDP. In M. Adler & M. Fleurbaey (Eds.), *The Oxford handbook of well-being and public policy*. Oxford University Press.

Schreyer, P., & Diewert, W. E. (2014). Household production, leisure and living standards. In D. W. Jorgenson, J. S. Landefeld, & P. Schreyer (Eds.), *Measuring economic sustainability and progress*. NBER Book Series Studies in Income and Wealth. http://www.nber.org/chapters/c12826

Sefton, J. A., & Weale, M. R. (2006). The concept of income in a general equilibrium. *Review of Economic Studies, 73,* 219–249.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the commission on measurement of economic performance and social progress*. https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report
UNDP. (2020). *Human development index*. http://hdr.undp.org/en/content/human-development-index-hdi
United Nations, Our Common Future, Report of the World Commission on Environment and Development. (1987). https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf
United Nations, European Union, Food and Agriculture Organization of the United Nations, Organisation for Economic Co-operation and Development, World Bank Group. (2014). *System of environmental-economic accounting 2012—Experimental ecosystem accounting*. https://seea.un.org/sites/seea.un.org/files/seea_eea_final_en_1.pdf
United Nations Statistical Commission. (2021). *System of environmental-economic accounting-ecosystem accounting: Final draft prepared by the committee of experts on environmental-economic accounting*. https://unstats.un.org/unsd/statcom/52nd-session/documents/BG-3f-SEEA-EA_Final_draft-E.pdf
Usher, D. (1976). The measurement of real income. *The Review of Income and Wealth, 22*(4), 305–3029.
Van de Ven, P., Zwijnenburg, J., & De Queljoe, M. (2018). *Including unpaid household activities: An estimate of its impact on macro-economic indicators in the G7 economies and the way forward* (OECD Statistics Working Papers 2018/04). OECD Publishing. http://dx.doi.org/10.1787/bc9d30dc-en
Vanoli, A. (1995). Reflections on environmental accounting issues. *Review of Income and Wealth, 41*(2) (June) 113–137.
Weitzman, M. L. (1976). On the welfare significance of national product in a dynamic economy. *The Quarterly Journal of Economics, 90,* 156–162.
Willig, R. (1981). Social welfare dominance. *American Economic Review, 71*(2) (May), 200–204.
World Bank. (2018). *The changing wealth of nations 2018*. The World Bank Publishing. https://openknowledge.worldbank.org/handle/10986/29001
Zwijnenburg, J., Bournot, S., & Giovannelli, F. (2017). *Expert group on disparities in a national accounts framework: Results from the 2015 exercise* (OECD Statistics Working Papers, No. 2016/10). OECD Publishing. https://doi.org/10.1787/2daa921e-en
Zwijnenburg, J., Bournot, S., Grahn, D., & Guidetti, E. (2021). *Distribution of household income, consumption and saving in line with national accounts: Methodology and results from the 2020 collection round* (OECD Statistics Working Papers).

CHAPTER 10

Hedonic Models and House Price Index Numbers

Robert J. Hill and Alicia N. Rambaldi

Introduction

This chapter provides a review of the literature on hedonic methods for constructing property price indices for residential housing. A number of methods for constructing price indices are based on the use and estimation of a hedonic regression, and thus are referred to in general as hedonic methods. However, there is a fundamental difference between methods that compute indices directly from the estimated parameters of the hedonic regression—the time-dummy method (section "Time-dummy Method") and the rolling-time-dummy method (section "Rolling Time–Dummy Method")—and those that compute indices from the imputed

R. J. Hill (✉)
Department of Economics, University of Graz, Graz, Austria
e-mail: robert.hill@uni-graz.at

A. N. Rambaldi
School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, St Lucia, QLD, Australia
e-mail: a.rambaldi@uq.edu.au

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_10

prices obtained from a hedonic regression model—the average characteristics method (section Average Characteristic Method), the hedonic imputation method (section "Hedonic Imputation Method") and the repricing method (section "Repricing Method").

Indices that are computed directly from parameter estimates have the advantage of readily providing standard errors; however, any biases due to model specification, such as omitted variables in the regression, are carried to the computed indices. Thus, careful model specification is required. Alternatively, indices that are computed from imputed prices require good prediction performance from the hedonic model. The concern is less on which individual variables to include in the model or whether there is collinearity. A well-performing imputation model can include a number variables that cover key predictors of property prices, such as location (see section "Controlling for Location"), and controls for land and dwelling characteristics that jointly explain the movements and distribution of property prices. Typical sample sizes are large enough so that degrees of freedom are not a concern, and collinearity will not affect the computed index since it does not affect predictions (i.e. imputations). Hedonic imputation methods were signalled as the preferred alternative in the Handbook on Residential Property Price Indices (HRPPI) (European Commission, Eurostat, OECD, and World Bank, 2013), which has been the most comprehensive compendium to date on methodology to construct residential property price indices. However, the field has moved on since the HRPPI was published. Many National Statistical Institutes (NSIs) are now using hedonic methods in their official indices that were not discussed in the HRPPI.

In section "Hedonic Methods for Constructing House Price Indices", we provide a comprehensive review of hedonic price index formulas building from earlier reviews and current practice at NSIs. This section also provides a systematic review of parametric and non-parametric alternatives for controlling for the location of the property in hedonic models used to construct price indices.

The hedonic approach is appealing in its flexibility. In particular, it can be extended to address other questions beyond basic index construction. A notable example is considered in section "Constructing Separate Price Indices for Land and Structures". This section shows how hedonic methods can be used to construct separate price indices for land and structure. Section "Extensions" reviews two other recent developments in the hedonic literature. The first is the computation of price indices at higher frequencies. The second is the construction of hedonic price indices for the whole housing stock rather than just for properties that have sold recently. Section "Conclusion" concludes.

Hedonic Methods for Constructing House Price Indices

This section is divided into two subsections. The first presents some of the hedonic methods that were covered in the HRPPI as they provide the basic framework to review a number of methods that have gained popularity since the HRPPI was published. In addition, one method—the repricing method—that was not discussed in the HRPPI is also considered. The second subsection considers ways of controlling for location effects in hedonic models.

Indices Covered in Chapter 5 of Eurostat's HRPPI

Time-Dummy Method

The time-dummy method estimates a single semi-log hedonic model as follows:

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + \varepsilon_h, \tag{10.1}$$

where h indexes all the housing transactions between periods b and t, p_h is the transaction price of property h, c indexes the set of available characteristics of the transacted properties, and ε is an identically, independently distributed error term with mean zero. The characteristics of the properties are given by $z_{c,h}$, while $d_{s,h}$ are dummy variables that equal 1 when a property was traded in period s, and zero otherwise.

The price index for period t relative to the base period b is then calculated as follows:

$$\frac{P_t}{P_b} = \exp(\hat{\delta}_t) \tag{10.2}$$

where $\hat{\delta}$ denotes the least squares estimate of δ.

The time-dummy method has three main attractions. First, it is relatively simple to use. Second, given it uses the full dataset it does not need as much data per period as other hedonic methods. Third, it provides standard errors on the estimated price indices.

The time-dummy method has two main weaknesses. First, the shadow prices can become stale, not reflecting the current state of the market when the hedonic model is estimated over many years. Second, whenever a new period is added to the dataset and the hedonic model re-estimated, all the price indices change.

Average Characteristic Method

The average characteristics method and the hedonic imputation method both begin by estimating the following semi-log hedonic model separately for each period. For example, for periods $t-1$ and t, the regression model takes the following forms[1]:

$$\ln p_{t-1,h} = \sum_{c=1}^{C} \beta_{t-1,c} z_{t-1,c,h} + \varepsilon_{t-1,h}, \tag{10.3}$$

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c} z_{t,c,h} + \varepsilon_{t,h}, \tag{10.4}$$

where h indexes the property transactions in period t, $p_{t,h}$ the transaction price, and $z_{t,h,c}$ is the level of characteristic c in dwelling h. No time dummies are included. The estimated shadow prices on the characteristics, $\beta_{t,c}$, are specific to period t and are updated every period.

A reference period is selected and an average basket of characteristics constructed for this period. This average basket of characteristics can be interpreted as an average property. The hedonic price index simply measures the change in the imputed price of this average property over time. A price index between periods $t - 1$ and t can now be calculated using the average property of period $t - 1$ (denoted by $\bar{z}_{t-1}$) as the reference:

$$\frac{\mathrm{P}_t}{\mathrm{P}_{t-1}} = \frac{\exp\left(\sum_{c=1}^{C} \beta_{t,c} \bar{z}_{t-1,c}\right)}{\exp\left(\sum_{c=1}^{C} \beta_{t-1,c} \bar{z}_{t-1,c}\right)} = \mathrm{P}^{\mathrm{L}}_{t-1,t}, \tag{10.5}$$

[1] This section draws extensively on Hill et al. (2018).

where $P^{L}_{t-1,t}$ denotes a Laspeyres-type price index between periods $t-1$ and t.

Alternatively, the average property of period t could be used as the reference as follows:

$$\frac{P_t}{P_{t-1}} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{t,c}\bar{z}_{t,c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{t-1,c}\bar{z}_{t,c}\right)} = P^{P}_{t-1,t}, \tag{10.6}$$

where $P^{P}_{t-1,t}$ denotes a Paasche-type price index. The terms $\bar{z}_{t-1,c}$ and $\bar{z}_{t,c}$ in Eqs. (10.5 and 10.6) denote the average baskets of characteristics of periods $t-1$ and t.

$$\bar{z}_{t-1,c} = \frac{1}{H_{t-1}} \sum_{h=1}^{H_{t-1}} z_{t-1,h,c}, \quad \bar{z}_{t,c} = \frac{1}{H_t} \sum_{h=1}^{H_t} z_{t,h,c}.$$

If one wants to treat both periods symmetrically, this can be done by taking the geometric mean of $P^{L}_{t-1,t}$ and $P^{P}_{t-1,t}$.

Each period the average characteristic basket is updated. Focusing on the Laspeyres case, relative to the base period b, the price index for period t is calculated as follows:

$$\frac{P_t}{P_b} = P^{L}_{b,b+1} \times P^{L}_{b+1,b+2} \times \cdots \times P^{L}_{t-1,t}.$$

The average characteristics method is still relatively simple and more market relevant than the time-dummy method in that the characteristic shadow prices are continually updated. However, estimating a separate hedonic model for each period can be problematic for smaller datasets. A second concern relates to the definition and interpretation of the average property. In particular, characteristics that take the form of dummy variables are probably best allocated fractionally to each category in proportion to the frequency in which they are observed.

Hedonic Imputation Method

The hedonic imputation method can be viewed as an extended version of the average characteristics method. Under certain conditions, as is shown below, the two methods are equivalent.

The underlying rationale of the hedonic imputation method is to use the hedonic model to impute missing prices so as then to allow standard price index formulas to be used.

Again we begin by estimating separate hedonic models for each period, as in Eqs. (10.3 and 10.4). Geometric-Laspeyres and geometric-Paasche-type formulas can now be computed as follows[2]:

$$\text{Geometric Laspeyres (GL)}: \frac{P_t}{P_{t-1}} = \left[\prod_{h=1}^{H_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})}\right]^{1/H_{t-1}}; \quad (10.7)$$

$$\text{Geometric Paasche (GP)}: \frac{P_t}{P_{t-1}} = \left[\prod_{h=1}^{H_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1,h}(z_{t,h})}\right]^{1/H_t}, \quad (10.8)$$

where $\hat{p}_{t,h}(z_{t-1,h})$ in Eq. (10.7) represents the predicted price in period t of a property with characteristic vector $z_{t-1,h}$ obtained from the hedonic model of period t, while $\hat{p}_{t-1,h}(z_{t-1,h})$ denotes the predicted price of the same property in period $t-1$ obtained from the hedonic model of period $t-1$. The terms in (10.8) have analogous interpretations.

Here we consider only double imputation indices. This means that both the numerator and denominator in each price relative is imputed. By contrast, single imputation imputes only the numerator for GL and only the denominator for GP (see de Haan, 2004; Hill & Melser, 2018; Pakes, 2003; Silver & Heravi, 2001). Double imputation is generally preferred since it partially controls for omitted variables in each price relative (see Hill & Melser, 2018).

Taking the geometric mean of GL and GP we obtain a Törnqvist-type index:

$$\text{Törnqvist}: \frac{P_t}{P_{t-1}} = \left\{\left[\prod_{h=1}^{H_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})}\right]^{1/H_{t-1}} \left[\prod_{h=1}^{H_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1,h}(z_{t,h})}\right]^{1/H_t}\right\}^{1/2}. \quad (10.9)$$

[2] Again this section draws extensively on Hill et al. (2018).

When the underlying hedonic model is semi-log, GL, GP and Törnqvist hedonic imputation indices can likewise be represented as average characteristic methods as follows (Hill & Melser, 2018):

$$\text{GL}: \left[\prod_{h=1}^{\text{H}_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})}\right]^{1/\text{H}_{t-1}} = \frac{\exp\left(\sum_{c=1}^{\text{C}} \hat{\beta}_{t,c}\bar{z}_{t-1,c}\right)}{\exp\left(\sum_{c=1}^{\text{C}} \hat{\beta}_{t-1,c}\bar{z}_{t-1,c}\right)} = \text{P}^{\text{L}}_{t-1,t}; \tag{10.10}$$

$$\text{GP}: \frac{\text{P}_t}{\text{P}_{t-1}} = \left[\prod_{h=1}^{\text{H}_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1,h}(z_{t,h})}\right]^{1/\text{H}_t} = \frac{\exp\left(\sum_{c=1}^{\text{C}} \hat{\beta}_{t,c}\bar{z}_{t,c}\right)}{\exp\left(\sum_{c=1}^{\text{C}} \hat{\beta}_{t-1,c}\bar{z}_{t,c}\right)} = \text{P}^{\text{P}}_{t-1,t}; \tag{10.11}$$

Törnqvist :

$$\left\{\left[\prod_{h=1}^{\text{H}_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})}\right]^{1/\text{H}_{t-1}} \left[\prod_{h=1}^{\text{H}_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1,h}(z_{t,h})}\right]^{1/\text{H}_t}\right\}^{1/2}$$
$$= \left\{\frac{\exp\left[\sum_{c=1}^{\text{C}} \hat{\beta}_{t,c}(\bar{z}_{t-1,c} + \bar{z}_{t,c})\right]}{\exp\left[\sum_{c=1}^{\text{C}} \hat{\beta}_{t-1,c}(\bar{z}_{t-1,c} + \bar{z}_{t,c})\right]}\right\}^{1/2}$$
$$= \left(\text{P}^{\text{L}}_{t-1,t} \times \text{P}^{\text{P}}_{t-1,t}\right)^{1/2}. \tag{10.12}$$

This duality between the average characteristics and hedonic imputation methods breaks down when the functional form of the hedonic model is not semi-log.

Rolling-Time-Dummy Method

The rolling-time-dummy (RTD) method estimates a time-dummy hedonic model on a rolling window of time periods (see O'Hanlon, 2011; Shimizu et al., 2010). Each time a new period of data becomes available, the rolling window is moved forward one period and the hedonic model re-estimated.

Price indices are derived from the estimated coefficients on the time dummies in the same way as with the time-dummy method except that each time the hedonic model is estimated we are only interested in the coefficient on the last period in the rolling window.[3]

Here we denote the first period in the window as period t. A semi-log hedonic model is estimated with a $k+1$ period window as follows:[4]

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{c,h} + \sum_{s=t+1}^{t+k} \delta_s d_{s,h} + \varepsilon_h, \tag{10.13}$$

where h indexes the housing transactions that occur within the rolling window. The set of available characteristics is indexed by c. The transaction price of property h is denoted by p_h, the property characteristics by $z_{c,h}$, and time-dummy variables capturing the period in which property h is sold by $d_{s,h}$. Finally, ε is a random error term with mean zero.

The change in the price index from period $t+k-1$ to period $t+k$ is then calculated as follows:

$$\frac{P_{t+k}}{P_{t+k-1}} = \frac{\exp\left(\hat{\delta}^t_{t+k}\right)}{\exp\left(\hat{\delta}^t_{t+k-1}\right)}, \tag{10.14}$$

where $\hat{\delta}$ denotes the least squares estimate of δ. The superscript t indicates that the estimated δ coefficient was obtained from the hedonic model with period t as the base (i.e. $P_t = 1$). This hedonic model is used only

[3] More sophisticated versions of the RTD method are developed in Hill et al. (2021).

[4] This section draws extensively on Hill et al. (2021).

to compute the change in house prices from period $t + k - 1$ to $t + k$. The window is then rolled forward one period and the hedonic model re-estimated. The price index comparing periods $t + k$ and $t + k + 1$ is now computed as follows:

$$\frac{P_{t+k+1}}{P_{t+k}} = \frac{\exp\left(\hat{\delta}_{t+k+1}^{t+1}\right)}{\exp\left(\hat{\delta}_{t+k}^{t+1}\right)}. \tag{10.15}$$

In Eq. (10.15) the base period in the hedonic model is now $t + 1$. Over multiple periods the price index is computed by chaining as follows:

$$\frac{P_{t+k+1}}{P_t} = \left[\frac{\exp(\hat{\delta}_{t+1}^{t-k})}{\exp(\hat{\delta}_{t}^{t-k})}\right] \left[\frac{\exp(\hat{\delta}_{t+2}^{t-k+1})}{\exp(\hat{\delta}_{t+1}^{t-k+1})}\right] \times \cdots \times \left[\frac{\exp(\hat{\delta}_{t+k+1}^{t+1})}{\exp(\hat{\delta}_{t+k}^{t+1})}\right]. \tag{10.16}$$

Unlike the time-dummy method, RTD price indices are never revised as new periods of data become available.

The RTD method is used by some European countries (Croatia, Cyprus, France, Ireland and Portugal) to compute their official house price indices (Hill et al., 2018). In addition, Japan's Official Property Price Index since 2012 uses RTD (Real Estate and Construction Economy Bureau, 2020). Japan has recently decided likewise to compute its commercial property price indices using the RTD method (see Shimizu & Diewert, 2019). Brunei Darussalam (see https://www.ambd.gov.bn/Site%20Assets%20%20News/RPPI-Technical-Notes.pdf), Peru and Thailand (see https://www.bot.or.th/App/BTWS_STAT/statistics/DownloadFile.aspx?file=EC_EI_008_S2_ENG.PDF) also use RTD, and Indonesia is about to start using it (see Rachman, 2019).

Given its popularity, it is perhaps surprising that RTD was not discussed in Chapter 5 of the HRPPI (on hedonic regression methods), except for the special case of a two-quarter rolling window—sometimes also referred to as the adjacent period method (see, e.g., Triplett, 2004). The RTD method was, however, discussed in Chapters 8 and 12, where it is recommended. The main reason RTD was excluded from Chapter 5 was probably because in 2013 it was still quite new. Also, RTD is a variant on the time-dummy method.

The reason RTD is now widely used is due to its attractive features. The method allows the index provider to choose the window length. This involves a trade-off. A longer window allows more data to be used each time the hedonic model is estimated, increasing the efficiency of the parameter estimates. By contrast, a shorter window increases the market relevance of the estimated shadow prices. Hence, in general bigger countries should choose shorter windows than smaller countries. The official house price index for France, for example, has a two-quarter rolling window, while the house price indices of Croatia and Cyprus have four-quarter rolling windows. This flexibility, combined with its simplicity and non-revisability, explains why the RTD method is becoming increasingly popular in recent years.

Repricing Method

The repricing method is used by Austria, Finland, Hungary, Italy, Latvia, Luxembourg, Norway and Slovenia to compute their official house price indices.[5] The repricing method, which is related to the average characteristics method, estimates a semi-log hedonic model using only the data of the base year b. The hedonic model can be written as follows:

$$\ln p_{b,h} = \sum_{c=1}^{C} \beta_{b,c} z_{b,h,c} + \varepsilon_{b,h}, \tag{10.17}$$

where h denotes a property sold in year b, $c = 1, \ldots, C$ indexes the characteristics of properties available in the dataset (such as floor area or number of bedrooms), and ε is a random error term.

The repricing price index formula divides a quality-unadjusted price index (QUPI) by a quality-adjustment factor (QAF). The QUPI is the ratio of the geometric mean prices in both periods $t-1$ and t, computed as follows:

$$\text{QUPI}_{t-1,t} = \frac{\tilde{p}_t}{\tilde{p}_{t-1}}, \tag{10.18}$$

[5] This section draws extensively on Hill et al. (2018).

where $\tilde{p}_{t-1}$ and $\tilde{p}_t$ denote the geometric mean price of properties sold in periods $t-1$ and t, respectively.

$$\tilde{p}_{t-1} = \prod_{h}^{\mathrm{H}_{t-1}} \left(p_{t-1,h}\right)^{1/\mathrm{H}_{t-1}}, \; \tilde{p}_t = \prod_{h}^{\mathrm{H}_t} \left(p_{t,h}\right)^{1/\mathrm{H}_t}. \tag{10.19}$$

In Eq. (10.19), H_{t-1} and H_t denote the number of properties sold in periods $t-1$ and t.

The quality-adjustment factor (QAF) uses the characteristic shadow prices $\hat{\beta}_b$ of year b to compare the cost of buying the average properties of periods $t-1$ and t as follows:

$$\mathrm{QAF}_{t-1,t} = \frac{\exp\left(\sum_{c=1}^{\mathrm{C}} \hat{\beta}_{b,c}\bar{z}_{t,c}\right)}{\exp\left(\sum_{c=1}^{\mathrm{C}} \hat{\beta}_{b,c}\bar{z}_{t-1,c}\right)}. \tag{10.20}$$

In Eq. (10.20),

$$\bar{z}_{t-1,c} = \frac{1}{\mathrm{H}_{t-1}} \sum_{h=1}^{\mathrm{H}_{t-1}} z_{t-1,h,c}, \quad \bar{z}_{t,c} = \frac{1}{\mathrm{H}_t} \sum_{h=1}^{\mathrm{H}_t} z_{t,h,c},$$

denote the average basket of characteristics of periods $t-1$ and t.

The repricing price index is obtained by dividing the quality-unadjusted index (QUPI) in Eq. (10.18) by the quality-adjustment factor (QAF) in Eq. (10.20) as follows:

$$\frac{P_t}{P_{(t-1)}} = \frac{QUPI_{t-1,t}}{QAF_{t-1,t}} = \frac{\tilde{p}_t}{\tilde{p}_{t-1}} \Bigg/ \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{b,c}\bar{z}_{t,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{b,c}\bar{z}_{(t-1),c})} \tag{10.21}$$

One attractive feature of the repricing method is that the characteristic shadow prices can be calculated using a full year of data, even when the index itself is being calculated on a quarterly basis. By contrast a quarterly average characteristics index computes the shadow prices each time using only one quarter's data, which can be problematic for smaller countries with fewer transactions per quarter.

Another interesting feature of the repricing method is that it only requires one set of shadow prices. However, failure to update the base period shadow prices can cause drift in the index.

The repricing method was not discussed in the HRPPI, and yet it is the most widely used method in Europe. The reason is because the repricing method was not well known before it received a strong recommendation in an early version of a Eurostat report on the treatment of owner-occupied housing (OOH) in the Harmonized Index of Consumer Prices (HICP). This report has been through a number of drafts—for example version 4 was available online from 2015 (Eurostat, 2015). The chapter on house price indices in this and earlier drafts was written completely independently from the HRPPI. A later draft from 2017 (also available online) includes the RTD method and does not endorse the repricing method.

National Statistical Institutes (NSIs) at the time of the earlier drafts were under pressure from Eurostat to start computing official house price indices (if they were not already doing so). A number of NSIs turned to the OOH Manual rather than the HRPPI for guidance and hence decided to use the repricing method. As was noted above, the repricing method is fine as long as the reference hedonic model is updated every year (as it is in Italy and Luxembourg). However, some European countries using the repricing method have not been updating their reference hedonic model as often as maybe they should.

Controlling for Location

Constructing a price index is intrinsically about the temporal dimension of property prices. However, these prices also show complex spatio-temporal relationships (for a recent review see Teye and Ahelegbey [2017] and their many references). The spatial dimension of this relationship is directly related to the physical location of properties. A property is an asset bundle composed of land and structure. Prices are determined by the interaction of the characteristics of these two assets. Location is a characteristic of the land component. The structure can be demolished and replaced, but the land will stay in the same physical location (see section "Constructing Separate Price Indices for Land and Structures"). In this section, we concentrate our attention on how location is measured and incorporated when constructing property price indices.

It is evident from the extensive review in the HRPPI European Commission, Eurostat, OECD, and World Bank (2013) and Hill (2013), that location has been a key consideration in the price index literature for a long time. All methodologies used to construct price indices for residential housing have attempted to incorporate or control for location in some form. Indices using stratification or mixed adjustment approaches constructed by the Australian Bureau of Statistics (ABS, 2006) use clustering of locations within cities. The standard repeat-sales method compares pairs of sales of the same address which provides a micro location control. More recent extensions of the hybrid repeat-sales/hedonic model (Case & Quigley, 1991; Hill et al., 1991) have been proposed which include a nearest-neighbour estimator to control for location (Gunternmann et al., 2016).

Hedonic regression-based methods can control for location in a number of ways. Location can be assumed to explain the behaviour of the mean of (log) price, the variance of (log) price or both.[6]

Approaches that assume location explains the mean of (log) price add parametric or non-parametric terms to the regression function. These approaches are discussed in section "Controlling for Location Dependence in the Hedonic Log-Price Function", and include post(zip)code dummy variables, distances to points of interest, a spatial price lag term and non-parametric approaches. The spatial error model, on the other hand, controls for location as explaining the variance of (log) prices (section "Controlling for Location Dependence in the Variance of the log-Price Function"). It is also possible to control for the effect of location on prices via both the mean and the variance of the hedonic model—presented in section "Controlling for Location Dependence in the Mean and the Variance of Log-Prices".

Controlling for Location Dependence in the Hedonic Log-Price Function

Including post(zip)code (neighbourhood) dummy intercepts

This is the traditional control for location that has been used in the price index literature. It is a parametric approach which consists of including a set of intercept dummy shifts into the regression model. To control for

[6] This discussion applies to Appraisal-Based Methods as they use a hedonic imputation model (see Chapter 7 of the HRPPI European Commission, Eurostat, OECD, and World Bank [2013]).

location using post(zip)code or neighbourhood dummies, a term is added to the hedonic regressions. For instance, the term Eq. (10.22) is added to models such as to those presented in Eqs. (10.1, 10.3, 10.4, and 10.13) or Eq. (10.17),

$$\sum_{l=1}^{PC} \lambda_l d_{lh} \tag{10.22}$$

where l indexes the postcodes of the market area from where the sample is drawn, d_{lh} is one if property h is in postcode l, and λ_l is a parameter that provides the size of the regression function shift associated with that post(zip)code location, and PC is the number of postcodes in the study area. This approach has been used extensively and has served to provide a base specification for comparing alternative approaches to controlling for location (Diewert & Shimizu, forthcoming; Hill & Scholz, 2018; Hill et al., 2021).

Using parametric and non-parametric functions of the coordinates
Geographical information systems (GIS) software allows for the expression of property addresses using coordinates (latitude, longitude). Currently, standard software packages such as R can work with coordinates and functions of coordinates embedded in statistical models. These advances provide a number of options for modelling. Eucledian distances between a given property and a point of interest (POI) or its nearest neighbours can be easily computed and used in the estimation of the hedonic model. Next parametric and non-parametric alternatives are presented.

Parametric options:

(a) **Distances to points of interest**

In this case, a set, L, of the hedonic characteristics in z_{ch} are regressors that control for location. This set is represented by the term Eq. (10.23) which is added to Eqs. (10.1, 10.3, 10.4, and 10.13) or Eq. (10.17). These regressors are either direct distances to major POI such us city centre, hospitals, shopping centres, schools, etc., or functions of these distances (e.g. inverse function).

$$\sum_{pi=1}^{L} \lambda_{pi} l_{pi,h} \tag{10.23}$$

Here $l_{pi,h}$ gives the distance in km(miles) to POI pi, and λ_{pi} is a parameter associated with the shadow price of that particular POI.

The use of these types of regressors is a common approach in the real estate and urban economics literature where the aim might be to estimate the willingness to pay associated with specific POIs (see, e.g., the review in D'Acci [2013]), in addition to using location to improve the prediction of (log) prices.

The price index literature has adopted the use of these type of regressors (Diewert & Shimizu, 2015; Rambaldi & Fletcher, 2014).

(b) **Spatial Lag model**

The spatial lag model (SLM) is a type of spatially dynamic model. It is an autoregressive model but on the spatial instead of the time dimension. A time-dummy hedonic regression Eq. (10.1), written in a spatially dynamic form is given by Eq. (10.24). A hedonic regression to construct hedonic imputated indices (such as 10.4) would be given by a specfication such as that in Eq. (10.25),

$$\ln p_h = \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + \varepsilon_h, \tag{10.24}$$

$$\ln p_{t,h} = \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_{t,c} z_{t,h,c} + \varepsilon_{t,h}, \tag{10.25}$$

where $|\rho_1| < 1$ is a spatial autoregressive parameter; $w_{hh} = 0$, since unit h is not its own neighbour; $\sum_{i=1}^{N} w_{hi} = 1$ for all i, the weighting matrix is row normalised; and $\varepsilon_{t,h}$ is assumed mean zero and uncorrelated.

The weights w_{hi} are based on the geographic distance between property h and property i, d_{hi}. Then,

$$\begin{cases} w_{hi} = \dfrac{d_{hi}}{\sum\limits_{i} d_{hi}} \text{ if } h \text{ and } i \text{ are neighbours} \\ w_{hi} = 0 \text{ if } h \text{ and } i \text{ are not neighbours} \end{cases} \tag{10.26}$$

There are some alternatives in how neighbours are defined. For example, "neighbours" can be defined either as a fixed number or as all of the first nearest neighbours (see Chapter 2, Kelejian & Piras, 2017).

To see how a **time-dummy index** can be computed from the SLM in Eq. (10.24), we use its reduced form in Eq. (10.27):

$$\ln p_h = \sum_{c=1}^{C} \beta_{c,w} z_{ch} + \sum_{s=b+1}^{t} \delta_{s,w} d_{sh} + \varepsilon_{h,w}, \tag{10.27}$$

where

$\beta_{c,w} = \frac{\beta_c}{(1-\rho \sum_{i=1}^{N} w_{hi})}$

$\delta_{s,w} = \frac{\delta_s}{(1-\rho \sum_{i=1}^{N} w_{hi})}$

The price index for period t relative to the base period b is then calculated as follows:

$$\frac{P_t}{P_b} = \exp(\hat{\delta}_{t,w}) \tag{10.28}$$

where $\hat{\delta}_{t,w}$ denotes the estimate of $\delta_{t,w}$ in Eq. (10.27).

To compute a **hedonic imputed price index** for period t, as in Eq. (10.9), the SLM in Eq. (10.25) can be written in its reduced form Eq. (10.29), and the required predictions, $\hat{p}_{t,h}(z_{t-1,h})$, $\hat{p}_{t-1,h}(z_{t-1,h})$; $\hat{p}_{t,h}(z_{t,h})$, and $\hat{p}_{t-1,h}(z_{t,h})$, obtained to compute the index.

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c,w} z_{t,h,c} + \varepsilon_{t,h,w}, \tag{10.29}$$

where $\beta_{t,c,w} = \frac{\beta_{t,c}}{(1-\rho \sum_{i=1}^{N} w_{hi})}$.

For further details on predictions from SLM, see Chapter 4 of Kelejian and Piras (2017).

Non-parametric options:

(a) **Splines and Spatial Coordinates**

The use of spatial coordinates to model location effects non-parametrically in the price index literature has been adopted in Hill and Scholz (2018), Diewert and Shimizu (forthcoming), and Hill et al. (2021).

Hill and Scholz (2018) proposed to use the semi-parametric model Eq. (10.30) to obtain the predictions required to compute a hedonic imputed index of the form in Eq. (10.9),

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c} z_{t,h,c} + g(\text{lat}_h, \text{long}_h)_t + \varepsilon_{t,h}, \tag{10.30}$$

This model is estimated using a penalised least squares approach. Hill and Scholz (2018) find the index computed from the predictions of this model at the quarterly frequency do not differ substantially from the index obtained using postcodes in place of the $g(.)$ function as in Eq. (10.22).

In a recent paper, Diewert and Shimizu (forthcoming) argue that the use of penalised least squares results in a smoothing method that fails the "smoothing invariance test" which implies the smooth series produced will change if a second round of smoothing is applied to the smoothed series originally obtained. They propose to use a modification of Colwell's (1998) spatial interpolation method. The modification can be viewed as a general non-parametric method for estimating a function of two variables. Their paper is concerned with constructing indices of the value of land and not of property prices (see section "Constructing Separate Price Indices for Land and Structures").

(b) **Geographically Weighted Regressions**

Geographically weighted regressions (GWR) are also in the family of modelling approaches that use functions of the latitude and longitude coordinates of the hth property, and they are non-parametric spatial models.

The GRW model is as follows,

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c,(\text{lat}_h,\text{long}_h)} z_{t,h,c} + \varepsilon_{t,h}, \tag{10.31}$$

The parameters $\beta_{t,c(\text{lat}_h,\text{long}_h)}$ are estimated using a Gaussian spatial kernel and the geographical distribution of the estimates are based on the Euclidean distance between observations. As with other kernel estimation techniques, there is a need to choose the bandwidth.

The observant reader would have noticed that the GRW model Eq. (10.31) is a non-parametric version of the SLM in Eq. (10.29). The use of a Gaussian kernel in this case would lead to the weight, w_{hi} associated with property i at location h being defined as:

$$w_{hi} = \exp\left[-1/2(\frac{d_{hi}}{b})^2\right] \tag{10.32}$$

where b is the bandwidth, and d_{hi} is defined as in Eq. (10.26). That is, the geographical distance between h and i.

Bidanset and Lombard (2014) compare the SML and the GRW in the context of mass appraisals for tax assessments using as a comparison metric the coefficient of dispersion (COD). Both models can provide geographically disaggregated estimates. Both dominate a geographically unaware model; however, neither is found to be the dominant over the other uniformly.

Constructing **hedonic imputed price indices** from the predictions of Eq. (10.31) follows the standard procedure of producing the four predictions required for the computation of the price index in Eq. (10.9). This is the same procedure as that stated for the SLM in Eq. (10.29). To obtain a **time-dummy hedonic index** would require a semi-parametric alternative where the hedonic regressors enter the regressions in a non-parametric form, while the time-dummy term is parametric. Bárcena et al. (2014) used a geographically weighted regression to study the distribution of prices, but then proposed to construct a price index from a semi-parametric model where the hedonic characteristics enter parametrically and a cubic spline function of time is then normalised to compute a non-parametric version of a time-dummy price index for the whole geographical area under study.

Controlling for Location Dependence in the Variance of the Log-Price Function

In this case, prices are assumed to be indirectly interrelated via spatially interrelated errors. This specification is then assuming that the covariance of prices is spatially dependent. A model for the computation of the time-dummy price index is given by Eq. (10.33), and one for the computation of hedonic imputed indices is given by Eq. (10.34),

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + u_h$$

$$u_h = \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_i + \varepsilon_h, \tag{10.33}$$

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_c z_{t,c,h} + u_{t,h}$$

$$u_{t,h} = \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_{t,i} + \varepsilon_{t,h}, \tag{10.34}$$

Note that both models can be written alternatively as Eqs. (10.35 and 10.36), respectively,

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + (1 - \rho_2 \sum_{i=1}^{N} w_{hi})^{-1} \varepsilon_h \tag{10.35}$$

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{ch} + u_h + (1 - \rho_2 \sum_{i=1}^{N} w_{hi})^{-1} \varepsilon_h \tag{10.36}$$

which shows why the error term in the hedonic model is not uncorrelated, and thus from first principles it follows that while OLS is a consistent estimator of the parameters of the model, the OLS computed standard errors will be biased. This model can be estimated by maximum likelihood (details are provided in Kelejian and Piras [2017]).

The computation of hedonic indices follows the standard approach, time-dummy indices from the estimated δ_s's and imputed indices from the predictions required for the computation of the price index in Eq. (10.9). The autoregressive spatial error structure of this model leads to at least two possible predictors, one is the standard predictor $\hat{p}_{t,h} = \exp[\sum_{c=1}^{C} \hat{\beta}_{t,c} z_{t,c,h}]$, the other is one that adds a correction due to the correlation induced from the spatial error lag in $u_{t,h}$ (the interested reader is directed to Chapter 4 of Kelejian and Piras [2017]).

Controlling for Location Dependence in the Mean and the Variance of Log-Prices

A general parametric model can be specified which includes both a spatial lag in the prices as well as in the error. Model Eq. (10.37) shows the specification to compute time-dummy hedonic indices, while model Eq. (10.38) shows the specification to compute hedonic imputed type indices.

$$\begin{aligned} \ln p_h &= \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + u_h \\ u_h &= \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_i + \varepsilon_h, \end{aligned} \tag{10.37}$$

$$\begin{aligned} \ln p_{t,h} &= \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_c z_{t,c,h} + u_{t,h} \\ u_{t,h} &= \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_{t,i} + \varepsilon_{t,h}, \end{aligned} \tag{10.38}$$

The estimation of these models is by maximum likelihood (details are provided in Kelejian and Piras [2017]).

Time-dummy indices can be computed from the estimated δ_s's and hedonic imputed price indices from the predictions following the standard procedure of producing the four predictions required for the computation of the price index in Eq. (10.9).

Empirical Feasibility

The previous subsections have provided a taxonomy of modelling approaches to compute both time-dummy and hedonic imputed price indices for residential housing that control for the dependence of prices on location. It is shown that all of the alternative specifications can be used to construct hedonic imputed property price indices. All alternatives are also easily implementable to construct time-dummy hedonic indices, except perhaps for the GWR model. Some authors have recently proposed to combine alternative models to improve price prediction (Oust et al., 2020), which as stated are inputs to hedonic imputed price indices. Importantly, there are packages in R, a toolbox in Matlab, and STATA routines that can estimate all or most of the above presented models making them feasible to practitioners everywhere.

Constructing Separate Price Indices for Land and Structures

Clapp (1980) first proposed a model for the level of property values that allowed for the notion of dividing the property into additive land and structure values. Bostic et al. (2008) proposed the concept of land leverage (the ratio of land value to overall property value) as an important indicator of residential property price dynamics and followed the additive formulation of land and structure.

The conceptual model is Eq. (10.39)

$$V = L + S \tag{10.39}$$

where V is the property value, L is the land value and S is the structure value.

The main issue faced by the modeller is that a standard hedonic regression cannot separate these two components. A log-linear specification cannot provide an additive decomposition. The regression must be linear. However, a standard linear regression with intercept (or time-dummies, or a time-varying intercept trend) and hedonic controls does not provide the required decomposition either. In this case, two mixed—land and structure—components are obtained: (1) overall market condition and (2) a hedonic quality adjustment. Intercept time dummies, or a trend, capture the macroeconomic conditions of the property market under study (combining price trends in both the land and the structure). The remaining part of the regression provides a combined quality-adjustment effect, where the individual estimates in the "hedonic quality adjustment" component, are measures of the marginal effects of additional units of land size (inside margin), location, bedrooms, bathrooms, etc. The realisation that it is not possible to separate the value of the land from that of the structure using standard regression estimates has led to a number of proposed alternatives that provide empirical identification strategies to separate the value of the land from that of the structure. Proposed approaches have included a non-linear systems approach, the use of exogenous information and imposing of asymmetric behaviours on the dynamics of the land and structure components.

Diewert (2007, Sect. 5.1) proposed to combine an additive and a log-linear model to be estimated as a non-linear system, which would provide price indices per square metre of structure and land. Diewert et al. (2011) explored a number of models and settled on a specification that used exogenous information to isolate the structure component, and thus providing identification of the land component. The model was then formalised in Diewert et al. (2015) and labelled "the builder's model". The approach to separating the value of land from structure is based on replacing the set of parameters associated with the structure by an official price index of new building construction and a non-linear adjustment due to the depreciation of the asset with age. This framework has been applied in Diewert et al. (2015) to data from the "Town of A" in the Netherlands, in Diewert and Shimizu (2015) in an application to Tokyo residential property, and in Diewert et al. (2017) in an application to British Columbia. By anchoring the model on an official price for new building construction, it is argued that the decomposition follows National Accounts principles and thus the estimates of land can be used in the computation of a country's productivity.

Rambaldi et al. (2010, 2016) proposed to approach the problem as the estimation of two unobserved components, where each component (land, structure) is uniquely mapped to a set of observable characteristics, and the behaviour of the components' prices is asymmetric. The underlying model is labelled "the valuer's model". The degree of asymmetry is determined by two bounded smoothing parameters which enter a modified Kalman filter algorithm. The land component is a function of land size and land location, and it is assumed to be the component that captures the largest proportion of price shocks in the market, an assumption that follows from earlier literature (Bostic et al., 2008). The structure component is a function of the structure's size (e.g. number of bedrooms, bathrooms, floor space, garages) and age,[7] and its value is assumed to be more stable as its movements follow the trends in local

[7] Building quality, e.g. building materials types, can also be added to the controls for the structure. However, data on these are less likely to be available. Empirically, these do

markets' wages, construction's inputs prices and depreciation. The implementation is simple as the model depends on three parameters that can be easily obtained. The first is the variance of the error term from a standard hedonic regression (obtained using least squares). The second is a pair of smoothing parameters which are bounded between zero and one and thus can be obtained by using a grid search. With estimates of these three parameters the algorithm to obtain the predictions of the value of the land and structure of each property, h, is just a set of formulae that does not require additional estimation. Rambaldi et al. (2016) compared their estimated price indices for land and structure for the "Town of A" in the Netherlands to those obtained by Diewert et al. (2015) to show they are not only comparable, but also smoother and can be computed at a monthly frequency even when the sample is small. Rambaldi and Tan (2019) computed land price indices for three regions within the Greater Melbourne (Australia) metropolitan area, and compared the index's predicted growth in land prices to those computed by the state of Victoria's Valuer-General (VGV). To illustrate we draw from Rambaldi and Tan's (2019) results.

Prior to 2019, revaluations from the VGV were run every two years and were part of the general valuation which also determines council

Table 10.1 Comparison of VGV land valuation versus model based Land Index

Region	*Revaluation*	*Benchmark (%)*[1]	*Model (%)*[2]	*Difference (%)*
Inner				
	2016–2018	27.40	29.30	−1.90
	2014–2016	30.54	22.71	7.83
Metro				
	2016–2018	29.00	35.60	−6.60
	2014–2016	33.81	27.90	5.91
Outer				
	2016–2018	45.71	36.48	9.23
	2014–2016	20.72	17.86	2.86

[1]VGV valuations are at 1 January of corresponding year (2016, 2018)
[2]Increase over the periods: 2013Q4:2015Q4, 2015Q4:2017Q4
Source Rambaldi and Tan (2019)

not seem to make a significant difference to the computed index. The key controls seem to be age and size of the structure.

Table 10.2 Comparison of VGV building cost index and model based structure index

Period	*VGV (residential construction)*[1]		*Model*[2]		
July–June	*Metropolitan*	*Regional*	*GM-Inner*	*GM-Metro*	*GM-Outer*
2008–2009	1.03	1.02	0.994	1.002	1.011
2009–2010	1.03	1.03	1.028	1.042	1.031
2010–2011	1.03	1.04	1.006	1.010	1.016
2011–2012	1.03	1.05	0.993	0.994	0.998
2012–2013	1	1.03	1.008	1.004	1.001
2013–2014	1.02	1.04	1.018	1.021	1.011
2014–2015	1	1	1.028	1.033	1.016
2015–2016	1.03	1.01	1.022	1.025	1.022
2016–2017	1.03	1.03	1.027	1.031	1.032
2017–2018	1.035	1.035	1.003	1.014	1.026

[1]These are reported for Metropolitan and Non-metro/Regional Victoria. The Metropolitan does not overlap exactly with what is defined as Greater Melbourne
Source https://www.dtf.vic.gov.au/financial-reporting-policy/valuer-general-building-indices
[2]The model produces disaggregated figures for three areas within the Greater Melbourne area (GM)
Source Rambaldi and Tan (2019)

rates. So, the valuation approach may have differed across Local Government Areas (LGAs) depending on the respective valuers' judgement. In early 2019 the VGV made available the revaluation outcomes for each LGA in the state of Victoria available on their website[8] since 2014.[9] These data contained the total site value (in $ amounts) for each LGA at a point in time. For example, the 2018 revaluation outcome determines the site value of properties as at 1 January 2018. These LGA site values were aggregated up to match the definition of inner, metro and outer regions of Greater Melbourne used by their model. From there, the biennial growth rate was calculated in line with that generated from the land value index (LVI[10]). Table 10.1 summarises the estimated revaluation outcome from the model's LVI, and compares to those from the VGV. They consider these results very encouraging given over this period

[8] https://www.propertyandlandtitles.vic.gov.au/valuation/council-valuations.

[9] 2018, 2016 and 2014 revaluation rounds.

[10] For example, an LGA in the inner area (SiteValue_inner_2018/SiteValue_inner_2016−1) × 100% compared against (LVI_2018/LVI_2016 − 1) × 100%.

the VGV only had oversight and there was a lack of standardisation. Table 10.2 compares the VGV's building cost index to that obtained from the model.

One of the motivations behind finding separate values for land and structure is to uncover the depreciation rate (impact of physical deterioration) of the stock of housing. The interested reader can consult Francke and van de Minne (2016) and Diewert et al. (2017) for a review of the literature and alternative approaches to the computation of the rate of depreciation of housing structures.

The price indices computed from the approach of Diewert and co-authors are of the time-dummy type as the index is based on a normalised set of time-period parameters that are estimated by the builder's model. The price indices computed from the approach proposed by Rambaldi and co-authors are of the hedonic imputation type. The model is used to compute the predictions of the price of land and structure for each sold property. Predictions of land prices are then used to compute formula (10.9), and similarly indices for the structure and the property (land + structure) can be obtained.

Extensions

Higher Frequency Indices

Traditionally property price indices have been computed at either the annual or quarterly frequency. Hedonic time-dummy-based indices typically fitted annual dummies to the model that then determined the frequency of the resulting index. Hedonic imputed price indices are computed from regressions where all parameters (intercept and those attached to the hedonic characteristics) change at each time period (year, quarter, month, etc.) (see section "Hedonic imputation method"). The price index literature achieved this requirement by re-estimating the regression each year or quarter. Depending on the sample size, it is feasible to follow this approach to compute hedonic imputed price indices at a monthly frequency. However, samples are not random and thus the composition of sales within a given month can have a large impact on the estimated parameters and predictions, and thus unduly influence the resulting index. This issue was raised by Rambaldi and Fletcher (2014), who proposed the use of time-varying parameter models to overcome the volatility induced by the composition of sales and varying sample sizes (due to issues such as seasonality of sales and periods of thin markets) when computing hedonic imputed price indices. Time-varying parameters build from the information from the previous and current periods producing a much smoother set of estimates and reducing the volatility of the imputations.

As more data are available, monthly price indices have become more common and until recently hailed as high frequency (see, e.g., Bárcena et al., 2014; Bourassa & Hoesli, 2017). Bollerslev et al. (2016) used an extended repeat-sales type model with data from ten major US cities to compute daily price indices. The model is estimated monthly, and then a moving-monthly window (i.e. it shifts the "month" by a day at a time) for the last month of the sample is used to produce a daily price index.

The first, to our knowledge, hedonic-based high frequency index is that by Hill et al. (2021). The model and index are computed at a weekly frequency using data for Sydney and a semi-parametric state-space model. Their model is a type of spatio-temporal specification, which have become popular in the real estate literature following the seminal work of Pace et al. (2000) (see, e.g., Chica-Olmo et al., 2019; Hawkins & Habib, 2018; Liu, 2013; Otto & Schmid, 2018; Teye & Ahelegbey, 2017). Parametric spatio-temporal models have been used to compute monthly hedonic imputed price indices for property prices by Rambaldi and Fletcher (2014) and for land prices by Rambaldi et al. (2016) and Rambaldi and Tan (2019).

An important finding of Hill et al. (2021) is that weekly indices are far more sensitive to the method of construction than those computed at a lower frequency such as quarterly. Hill and Scholz (2018) and Diewert and Shimizu (forthcoming) found hedonic imputed price indices obtained using postcode dummies to control for location in the hedonic model do not differ significantly from those obtained with models that use more sophisticated specifications, such as splines (see section "Controlling for Location" for a presentation of alternative methods). Using the same metric as that proposed in Hill and Scholz (2018) to compare indices—Index MSE(RS)—Hill et al. (2021) find the indices obtained by a spatio-temporal model produces significantly and uniformly superior predictions of price relatives (i.e. the building blocks of a price index) to those obtained with Hill and Scholz (2018)—GAM—and using postcode dummies in a time-varying parameter model (SS+PC) at monthly and weekly frequencies. This is shown in Tables 10.3 and 10.4, which have

Table 10.3 Model Prediction and Index Quality Comparison

	Model RMSPE				*Index MSE(RS)*	
	Sydney	*Harbour*	*Bondi beach*	*Blue Mountains*	*Weekly*	*Monthly*
Radius		5 Km	2.5 Km	30 Km		
GAM	0.1857	0.3136	0.3008	0.1260	0.0233	0.0245
SS + GAM	0.1775	0.3067	0.2954	0.1315	0.0102	0.0112
SS + PC	0.2088	0.3518	0.3239	0.1540	0.0246	0.0264
Sample	433202	13222	6950	19089		

Note The mean square prediction error of prices (RMSPE) are uniformly higher for the model with postcodes across all geographical alternatives. Similarly, the mean square error of the prediction of price relatives (MSE(RS)) are higher at both the SS + PC at both weekly and monthly frequency. The RMSPE is lowest for the SS + GAM model except in one case (the Blue Mountains) when GAM is the lowest. The SS + GAM is uniformly the lowest in MSE(RS) for both weekly and monthly frequencies

Reproduced from Hill et al. (2021)—Table 10.3

Table 10.4 *p*-values for H_0: $MSE(RS)_{M1} - MSE(RS)_{M2} = 0$

	Weekly	*Monthly*
SS + PC vs. SS + GAM	0.0000	0.0000
SS + PC vs. GAM	0.0483	0.0014
GAM vs. SS + GAM	0.0000	0.0000

Note These *p*-values imply that SS + GAM is highly significantly different from both SS + PC and GAM at both the weekly and monthly frequencies

Reproduced from Hill et al. (2021)—Table 10.4

been reproduced from Hill et al. (2021). Their proposed spatio-temporal model is labelled SS + GAM.

Measuring Price Changes for the Stock of Housing

The combination of data availability, mass-imputation techniques and high computing power provides an ideal environment to consider constructing price changes for the stock of housing. This issue was mentioned in the HRPPI (European Commission, Eurostat, OECD, and World Bank, 2013), Chapters 4 and 8. The HRPPI (European Commission, Eurostat, OECD, and World Bank, 2013) indicated that the use of stratification can approximate a stock-based residential property price index. Diewert et al. (2017) propose to use sales data over a reasonably long period of time to approximate the quantity (stock) of residential property. The construction of the relevant "stock" of housing is the key issue. The availability of administrative data would seem to be a promising path. Administrative land titles' data can provide the population of property by use (e.g. residential detach, attached, etc.). However, these would need to be linked to other datasets that capture renovations and improvements to provide a reasonable approximation of the stock at each point in time. Once a stock dataset of characteristics at the level of individual properties has been constructed, a hedonic model can be used to estimate prices for these properties each period. A hedonic imputation price index for the housing stock can then be computed. This is an area likely to see more research in the near future.

Conclusion

In recent years, the rolling-time dummy (RTD) and repricing methods have become popular with National Statistical Institutes (NSIs) for constructing official house price indices. Indeed most NSIs in Europe use one of these two methods. One reason for this is that both methods are better suited for use with smaller datasets than the average characteristics or hedonic imputation methods.

Location is typically controlled for in hedonic models using postcode/zipcode dummy variables. However, a number of more sophisticated methods are now available, particularly given the increasing availability of geo-coded longitudes and latitudes at the level of individual properties. While such methods are generally not currently being used by NSIs in their official indices, this could change in the future.

Another active area of research is the use of hedonic methods to construct separate price indices for land and structures. A key concern here is that house price indices may be upwardly biased if they fail to account for depreciation of the structures. Separating land from structure ensures that the resulting land price index is not distorted by depreciation.

There is growing demand for higher frequency (e.g. weekly) indices. In some cases, the housing datasets may not be large enough to easily accommodate say weekly indices. Recently, a number of approaches have been developed to allow more robust house price indices to be computed at higher frequencies and/or on smaller datasets.

Finally, progress is also being made on the construction of price indices for the stock of housing. The hedonic imputation method is ideal for this purpose as long as characteristic information is available on a sufficiently large portion of the housing stock.

In conclusion, the application of hedonic methods to the construction of house price indices is an active research area in which significant progress has been made in the last few years. This is helping to improve the accuracy of house price indices and broadening the range of indices that can be computed.

Acknowledgments We would like to thank Chihiro Shimizu, D.S. Prasada Rao and Nicholas Rohde for comments on an earlier version. Hill acknowledges financial support from the Austrian Research Promotion Agency (FFG), grant 10991131. Rambaldi acknowledges financial support from the Australian Research Council (ARC), grants DP120102124 and LP160101518.

References

Australian Bureau of Statistics. (2006). *A guide to the house price indexes* (Cat 6464.0).

Bárcena, M. J., Menéndez, P., Palacios, M. B., & Tusell, F. (2014). *A real-time property value index based on web data, data mining applications with R, Chapter 10*. Elsevier.

Bidanset, P. E., & Lombard, J. R. (2014). Evaluating spatial model accuracy in mass real estate appraisal: A comparison of geographically weighted regression and the spatial lag model. *School of Public Service Faculty Publications, 26*. https://digitalcommons.odu.edu/publicservice_pubs/26

Bollerslev, T., Patton, A. J., & Wang, W. (2016). Daily house price indices: Construction, modeling, and longer-run predictions. *Journal of Applied Econometrics, 31*, 1005–1025. https://doi.org/10.1002/jae.2471

Bostic, R. W., Longhofer, S. D., & Readfearn, C. L. (2007). Land leverage: Decomposing home price dynamics. *Real Estate Economics, 35*(2), 183–208.

Bourassa, S. C., & Hoesli, M. (2017). High-frequency house price indexes with scarce data. *Journal of Real Estate Literature, 25*(1), 207–220.

Case, K. E., & Quigley, J. M. (1991). The dynamics of real estate prices. *The Review of Economics and Statistics, 73*(1), 50–58.

Chica-Olmo, J., Cano-Guervos, R., & Chica-Rivas, M. (2019). Estimation of housing price variations using spatio-temporal data. *Sustainability, 11*, 1551. https://doi.org/10.3390/su11061551

Clapp, J. M. (1980). The elasticity of substitution for land: The effects of measurement errors. *Journal of Urban Economics, 8,* 255–263.

Colwell, P. F. (1998). A primer on piecewise parabolic multiple regression analysis via estimations of Chicago CBD land prices. *Journal of Real Estate Finance and Economics, 17*(1), 87–97.

D'Acci, L. (2013). Monetary, subjective and quantitative approaches to assess urban quality of life and pleasantness in cities (hedonic price, willingness-to-pay, positional value, life satisfaction, Isobenefit lines). *Social Indicators Research, 115,* 531–559. https://doi.org/10.1007/s11205-012-0221-7

de Haan, J. (2004). Direct and indirect time dummy approaches to hedonic price measurement. *Journal of Economic and Social Measurement, 29*(4), 427–443.

Diewert, W. E. (2007). *The Paris OECD-IMF workshop on real estate price indexes: Conclusions and future directions* (Discussion Paper 07-01). Department of Economics, University of British Columbia.

Diewert, W. E., de Haan, J., & Hendriks, R. (2011). The decomposition of a house price Index into land and structures components: A hedonic regression approach. *The Valuation Journal, 6,* 58–106.

Diewert, W. E., de Haan, J., & Hendriks, R. (2015). Hedonic regressions and the decomposition of a house price index into land and structure components. *Econometric Reviews, 34,* 106–126.

Diewert, W. E., Huang, N., & Burnett-Isaacs, K. (2017). *Alternative approaches for resale housing price indexes* (Discussion Paper 17-05). Vancouver School of Economics, The University of British Columbia.

Diewert, W. E., & Shimizu, C. (2015). Residential property price indexes for Tokyo. *Macroeconomic Dynamics 1*(8), 1–56. https://doi.org/10.1017/S1365100514000042

Diewert, W. E., & Shimizu, C. (Forthcoming). Residential property price indexes, spatial coordinates versus neighbourhood dummy variables. *Review of Income and Wealth*. https://doi.org/10.1111/roiw.12534

European Commission, Eurostat, OECD, and World Bank. (2013). *Handbook on Residential Property Price indices (RPPIs)*. Publications Office of the European Union. https://doi.org/10.2785/34007

Eurostat. (2015). *Detailed technical manual on owner-occupied housing for harmonised index of consumer prices (DF V4)*. Eurostat.

Francke, M. K., & van de Minne, A. M. (2016). Land, structure and depreciation. *Real Estate Economics*, 45(2), 1–37. https://doi.org/10.1111/1540-6229.12146

Guntermann, K. L., Liu, C., & Nowak, A. D. (2016, January–March). Price indexes for short horizons, thin markets or smaller cities. *The Journal of Real Estate Research, 38*(1), 93–128.

Hawkings, J., & Habib, K. N. (2018). Spatio-temporal hedonic price model to investigate the dynamics of housing prices in contexts of urban form and transportation services in Toronto. *Transportation Research Record, 2672*(6), 21–30. https://doi.org/10.1177/0361198118774153

Hill, R. C., Knight, J. R., & Sirmans, C. F. (1991). Estimating capital asset price indexes. *Review of Economics and Statistics, 79*(2), 226–233.

Hill, R. J. (2013). Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys, 27*(5), 879–914. https://doi.org/10.1111/j.1467-6419.2012.00731.x

Hill, R. J., & Melser, D. (2008, October). Hedonic imputation and the price index problem: An application to housing. *Economic Inquiry, 46*(4), 593–609. https://doi.org/10.1111/j.1465-7295.2007.00110.x

Hill, R. J., Rambaldi, A., & Scholz, M. (2021). Higher frequency hedonic house price indices: A state-space approach. *Empirical Economics, 61,* 417–441. https://doi.org/10.1007/s00181-020-01862-y

Hill, R. J., & Scholz, M. (2018). Can geospatial data improve house price indexes? A hedonic imputation approach with splines. *Review of Income and Wealth, 64*(4), 737–756. https://doi.org/10.1111/roiw.12303

Hill, R. J., Scholz, M., Shimizu, C., & Steurer, M. (2018). An evaluation of the methods used by European countries to compute their official house price indices. *Economie et Statistique / Economics and Statistics*, 500-501-502, 221–238. https://doi.org/10.24187/ecostat.2018.500t.1953

Hill, R. J., Scholz, M., Shimizu, C., & Steurer, M. (2021). Rolling-time-dummy house price indexes: Window length, linking and options for dealing with low transaction. *Journal of Official Statistics*, forthcoming.

Kelejian, H., & Piras, G. (2017). *Spatial econometrics*. Elsevier, Academic Press.

Liu, X. (2013). Spatial and temporal dependence in house price prediction. *Journal of Real Estate Finance and Economics, 47,* 341–369. https://doi.org/10.1007/s11146-011-9359-3

O'Hanlon, N. (2011). Constructing a national house price index for Ireland. *Journal of the Statistical and Social Inquiry Society of Ireland, 40,* 167–196.

Otto, P., & Schmid, W. (2018). Spatiotemporal analysis of German real-estate prices. *Annals of Regional Science, 60*, 41–72. https://doi.org/10.1007/s00168-016-0789-y

Oust, A., Hansen, S. N., & Pettrem, T. R. (2020). Combining property price predictions from repeat sales and spatially enhanced hedonic regressions. *Journal of Real Estate Finance and Economics, 61,* 183–207. https://doi.org/10.1007/s11146-019-09723-x

Pace, R. K., Barry, R., Gilley, O. W., & Sirmans, C. F. (2000). A method for spatial–temporal forecasting with an application to real estate prices. *International Journal of Forecasting, 16*, 229–246. https://doi.org/10.1016/S0169-2070(99)00047-3

Pakes, A. (2003). A reconsideration of hedonic price indices with an application to PC's. *American Economic Review, 93*(5), 1578–1596.

Rachman, A. N. (2019, February 20–22). *An alternative hedonic residential property price index for Indonesia using Big Data: The case of Jakarta*. Paper presented at the International Conference on Real Estate Statistics, Eurostat. https://www.crrem.eu/crrem-at-the-eurostat-international-conference-on-real-estate-statistics-2019/

Rambaldi, A. N., & Fletcher, C. S. (2014). Hedonic imputed property price indexes: The effects of econometric modeling choices. *Review of Income and Wealth, 60*, S423–S448. https://doi.org/10.1111/roiw.12143

Rambaldi, A. N., McAllister, R. R. J., & Fletcher, C. S. (2010). *Separating land from structure in property prices: A case study from Brisbane Australia*. School of Economics, The University of Queensland.

Rambaldi, A. N., McAllister, R. R. J., & Fletcher, C. S. (2016, August). *Decoupling land values in residential property prices: Smoothing methods for hedonic imputed price indices*. 34th IARIW General Conference. http://old.iariw.org/dresden/rambaldi.pdf

Rambaldi, A. N., & Tan, M. S. (2019). *Land value indices and the land leverage hypothesis in residential housing*. International Conference on Real Estate Statistics, EuroStat. https://ec.europa.eu/eurostat/about/opportunities/conferences

Real Estate and Construction Economy Bureau. (2020). *Methodology of JRPPI: Japan Residential Property Price Index*. Ministry of Land, Infrastructure, Transport and Tourism. https://www.mlit.go.jp/common/001360414.pdf

Shimizu, C., & Diewert, W. E. (2019). *Residential property price index in Japan: Discussion in methodology and data sources*. Paper presented at the International Conference on Real Estate Statistics, Eurostat. https://www.crrem.eu/crrem-at-the-eurostat-international-conference-on-real-estate-statistics-2019/

Shimizu, C., Takatsuji, H., Ono, H., & Nishimura, K. G. (2010). Structural and temporal changes in the housing market and hedonic housing price indices. *International Journal of Housing Markets and Analysis, 3*(4), 351–368.

Silver, M., & Heravi, S. (2001, April 2–6). *Quality adjustment, sample rotation and CPI practice: An experiment*. Presented at the Sixth Meeting of the International Working Group on Price Indices.

Teye, A. L., & Ahelegbey, D. F. (2017). Detecting spatial and temporal house price diffusion in the Netherlands: A Bayesian network approach. *Regional Science and Urban Economics*. https://doi.org/10.1016/j.regsciurbeco.2017.04.005

Triplett, J. E. (2004). *Handbook on hedonic indexes and quality adjustments in price indexes: Special application to information technology products* (STI Working Paper 2004/9). Directorate for Science, Technology and Industry, Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/science-and-technology/handbook-on-hedonic-indexes-and-quality-adjustments-in-price-indexes_643587187107

CHAPTER 11

Scanner Data, Elementary Price Indexes and the Chain Drift Problem

W. Erwin Diewert

Introduction

The *Consumer Price Index Manual*[1] recommended that the Fisher, Walsh or Törnqvist Theil price index be used as a *target month-to-month index* in a Consumer Price Index, provided that monthly price and expenditure data for the class of expenditures in scope were available. In recent years, retail chains in several countries (e.g., Australia, Canada, Japan, the Netherlands, Norway and Switzerland) have been willing to donate their

[1] See paragraph 22.63 in the ILO, Eurostat, IMF, OECD, UN and the World Bank (2004).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-981-19-2023-3_11.

W. E. Diewert (✉)
University of British Columbia, Vancouver, BC, Canada
e-mail: erwin.diewert@ubc.ca

University of New South Wales, Sydney, NSW, Australia

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_11

sales value and quantity sold information by detailed product to their national statistical agencies so it has become possible to calculate month-to-month superlative indexes for at least some strata of the country's Consumer Price Index.[2] However, the following issue arises: should the indexes fix a base month (for 12 or 13 months) and calculate Fisher fixed base indexes or should they calculate chained month-to-month indexes Fisher indexes? The 2004 *CPI Manual* offered the following advice on this choice in the chapter on seasonal commodities[3]:

- Determine the set of commodities that are present in the marketplace in both months of the comparison of prices between the two periods.
- For this maximum overlap set of commodities, calculate one of the three indexes recommended in previous chapters using the chain principle, i.e., calculate the chained Fisher, Walsh or Törnqvist Theil index.

The *CPI Manual* suggested the use of chained superlative indexes as a target index for the following three reasons[4]:

- The set of seasonal commodities which overlaps during two consecutive months is likely to be much larger than the set obtained by comparing the prices of any given month with a fixed base month (like January of a base year). Hence the comparisons made using chained indexes will be more comprehensive and accurate than those made using a fixed base.
- In many economies, on average 2 or 3% of price quotes disappear each month due to the introduction of new commodities and the disappearance of older ones. This rapid sample attrition means that fixed base indexes rapidly become unrepresentative, and hence, it

[2] Some countries may be able to obtain price and quantity data for individual products from third party data aggregators. This can be a cost-effective strategy for a statistical agency. In other cases, price and quantity data for regulated industries can be obtained from regulators.

[3] For more on the economic approach and the assumptions on consumer preferences that can justify month-to-month maximum overlap indexes, see Diewert (1999a, 51–56).

[4] See the ILO, Eurostat, IMF, OECD, UN and the World Bank (2004, 407).

seems preferable to use chained indexes that can more closely follow marketplace developments.

- If prices and quantities are trending relatively smoothly over time, chaining will reduce the spread between the Paasche and Laspeyres indexes.[5] Since these indexes provide reasonable bounds for true cost of living indexes, reducing the spread between these indexes will narrow the zone of uncertainty about the cost of living.

Thus the 2004 *Manual* recommended the use of chained Fisher, Walsh or Törnqvist Theil indexes as a target index concepts. But, as will be seen in the subsequent text, this advice does not always work out too well.

The problem with the above advice is the assumption of smooth trends in prices and quantities. Hill (1993, 388), drawing on the earlier research of Szulc (1983, 1987) and Hill (1988, 136–137), noted that it is not appropriate to use the chain system when prices oscillate or "bounce" to use Szulc's (1983, 548) term. This phenomenon can occur in the context of regular seasonal fluctuations or in the context of sales. The extent of the *price bouncing problem* or the problem of *chain drift* can be measured if we make use of the following test due to Walsh (1901, 389), (1921b, 540)[6]:

$$Multiperiod\ Identity\ Test: P\left(p^0, p^1, q^0, q^1\right) \times P\left(p^1, p^2, q^1, q^2\right) P\left(p^2, p^0, q^2, q^0\right) = 1$$

where $p^t \equiv [p_{t1}, ..., p_{tN}]$ and $q^t \equiv [q_{t1}, ..., q_{tN}]$ are the period t price and quantity vectors and p_{tn} and q_{tn} are the period t price and quantity for commodity n for $n = 1, \ldots, N$ in the class of commodities under consideration. $P\left(p^0, p^1, q^0, q^1\right)$ is a bilateral index number formula that

[5] See Diewert (1978, 895) and Hill (1988, 1993, 387–388). Chaining under these conditions will also reduce the spread between fixed base and chained indexes using P_F, P_W or P_T as the basic bilateral formula.

[6] Fisher (1922, 293) realized that the chained Carli, Laspeyres and Young indexes could be subject to upward chain drift but for his empirical example, there was no evidence of chain drift for the Fisher formula. However, Persons (1921, 110) came up with an empirical example where the Fisher index exhibited substantial downward chain drift. Frisch (1936, 9) seems to have been the first to use the term "chain drift." Both Frisch (1936, 8–9) and Persons (1928, 100–105) discussed and analyzed the chain drift problem. These indexes will be formally defined later in the chapter.

Table 11.1 Price and quantity data for two products for four periods

Period t	p_{t1}	p_{t2}	q_{t1}	q_{t2}
1	1.0	1.0	10	100
2	0.5	1.0	5000	100
3	1.0	1.0	1	100
4	1.0	1.0	10	100

is a function of the prices and quantities of periods 0 and 1. Thus price change is calculated over consecutive periods but an artificial final period is introduced as the final period where the prices and quantities revert back to the prices and quantities in the very first period. The test asks that the product of all of these price changes should equal unity. If prices have no definite trends but are simply bouncing up and down in a range, then the above test can be used to evaluate the amount of chain drift that occurs if chained indexes are used under these conditions. *Chain drift* occurs when an index does not return to unity when prices in the current period return to their levels in the base period.[7] Fixed base indexes that satisfy the time reversal test will satisfy Walsh's test and hence will not be subject to chain drift as long as the base period is not changed.

The *Manual* did not take into account how severe the chain drift problem could be in practice.[8] The problem is mostly caused by *sales* (i.e., highly discounted prices) of products.[9] An example will illustrate the problem.

Suppose that we are given the price and quantity data for two commodities for four periods. The data are listed in Table 11.1.[10]

The first commodity is subject to periodic sales (in period 2), when the price drops to ½ of its normal level of 1. In period 1, we have "normal" off sale demand for commodity 1 which is equal to 10 units. In period 2,

[7] See the ILO, Eurostat, IMF, OECD, UN and the World Bank (2004, 445).

[8] Szulc (1983, 1987) demonstrated how big the chain drift problem could be using chained Laspeyres indexes but the authors of the 2004 *Manual* did not realize that chain drift could also be a problem with chained *superlative* indexes.

[9] Pronounced fluctuations in the prices and quantities of seasonal commodities can also cause chain drift.

[10] This example is taken from Diewert (2012).

Table 11.2 Fixed base and chained Fisher, Törnqvist-Theil, Laspeyres and Paasche indexes

Period	$P_{F(FB)}$	$P_{L(FB)}$	$P_{P(FB)}$	$P_{F(CH)}$	$P_{T(CH)}$	$P_{L(CH)}$	$P_{P(CH)}$
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.698	0.955	0.510	0.698	0.694	0.955	0.510
3	1.000	1.000	1.000	0.979	0.972	1.872	0.512
4	1.000	1.000	1.000	0.979	0.972	1.872	0.512

the sale takes place and demand explodes to 5000 units.[11] In period 3, the commodity is off sale and the price is back to 1 but many shoppers have stocked up in the previous period so demand falls to only 1 unit. Finally in period 4, the commodity is off sale and we are back to the "normal" demand of 10 units. Commodity 2 exhibits no price or quantity change across periods: its price is 1 in all periods and the quantity sold is 100 units in each period. Note that the only thing that has happened going from period 3 to 4 is that the demand for commodity one has picked up from 1 unit to the "normal" level of 10 units. Also note that, conveniently, the period 4 data are exactly equal to the period 1 data so that for Walsh's test to be satisfied, the product of the period to period chain links must equal one.

Table 11.2 lists the fixed base Fisher, Laspeyres and Paasche price indexes, $P_{F(\text{FB})}$, $P_{L(\text{FB})}$ and $P_{P(\text{FB})}$ and as expected, they behave perfectly in period 4, returning to the period 1 level of 1. Then the chained Fisher, Törnqvist-Theil, Laspeyres and Paasche price indexes, $P_{F(\text{CH})}$, $P_{T(\text{CH})}$, $P_{L(\text{CH})}$ and $P_{P(\text{CH})}$ are listed. Obviously, the chained Laspeyres and Paasche indexes have chain drift bias that is extraordinary but what is interesting is that the chained Fisher has a 2% downward bias and the chained Törnqvist has a close to 3% downward bias.

What explains the results in the above table? The problem is this: when commodity one comes off sale and goes back to its regular price in period 3, *the corresponding quantity does not return to the level it had in period*

[11] This example is based on an actual example that used Dutch scanner data. When the price of a detergent product went on sale in the Netherlands at approximately one half of the regular price, the volume sold shot up approximately one thousand fold; see de Haan (2008, 15) and de Haan and van der Grient (2011). These papers brought home the magnitude of volume fluctuations due to sales and led Ivancic et al. (2009, 2011) to propose the use of rolling window multilateral indexes to mitigate the chain drift problem.

1: the period 3 demand is only 1 unit whereas the "normal" period 1 demand for commodity 1 was 10 units. It is only in period 4, that demand for commodity one recovers to the period 1 level. However, since prices are the same in periods 3 and 4, all of the chain links show no change (even though quantities are changing) and this is what causes the difficulties. If demand for commodity one in period 3 had immediately recovered to its "normal" period 1 level of 10, then there would be no chain drift problem.[12]

There are at least four possible real-time solutions to the chain drift problem:

- Use a fixed base index;
- Use a multilateral index[13];
- Use annual weights for a past year or
- Give up on the use of weights at the first stage of aggregation and simply use the Jevons index, which does not rely on representative weights.

There are two problems with the first solution: (i) the results depend asymmetrically on the choice of the base period and (ii) with new and disappearing products,[14] the base period prices and quantities may lose

[12] If the economic approach to index number theory is adopted, what causes chain drift in the above example is *inventory stocking behavior* on the part of households. The standard theory for the cost of living index implicitly assumes that all purchased goods are nondurable and used up in the period of purchase. In real life households can stockpile goods when they go on sale and it is this stockpiling phenomenon that leads to downward chain drift for a superlative index. For an example, where a chained superlative index has upward chain drift, see section "To Chain or Not to Chain". Feenstra and Shapiro (2003) also looked at the chain drift problem that was caused by sales and restocking dynamics. Their suggested solution to the chain drift problem was to use fixed base indexes which was also the advice of Persons (1921, 112).

[13] A multilateral price index compares average price levels over multiple periods. A bilateral price index compares price levels over two periods. Multilateral price indexes were originally applied in making *cross country* comparisons of prices. The use of multilateral indexes in the time series context dates back to Persons (1921) and Fisher (1922, 297–308), Gini (1931) and Balk (1980, 1981). Fisher (1922, 305) suggested taking the arithmetic average of the Fisher "star" indexes whereas Gini suggested taking the geometric mean of the star indexes. For additional material on multilateral indexes, see Diewert (1988, 1999b), Balk (1996, 2008) and Diewert and Fox (2020).

[14] We use the term "products" as meaning "goods and services."

their representativeness; i.e., over long periods of time, matching products becomes very difficult.[15]

A problem with the second solution is that as an extra period of data becomes available, the indexes may have to be recomputed. This is not a major problem. A solution to this problem is to use a rolling window of observations and use the results of the current window to update the index to the current period. This methodology was suggested by Ivancic et al. (2009, 2011) and is being used by the Australian Bureau of Statistics (2016). There is the problem of deciding exactly how to link the results of the current rolling window to the indexes generated by the previous rolling window but again, this is not a major problem.[16] However, it is possible to solve these linking problems by making use of a different class of multilateral methods, namely methods that rely on linking the data of the current period with a prior period that has the most similar structure of relative prices. This new class of multilateral methods will be explained in sections "Linking Based on Relative Price Similarity" and "Linking Based on Relative Price and Quantity Similarity".

The problem with the third possible solution is that the use of annual weights will inevitably result in some substitution bias, usually in the range of 0.15 to 0.40 percentage points per year.[17]

The problem with the fourth possible solution is that the use of an index that does not use quantity or expenditure weights will give equal

[15] Persons (1928, 99–100) has an excellent discussion on the difficulties of matching products over time.

[16] Ivancic et al. (2009, 2011) suggested that the movement of the rolling window indexes for the last two periods in the new window be linked to the last index value generated by the previous window. However, Krsinich (2016) suggested that the movement of the indexes generated by the new window be linked to the previous window index value for the second period in the previous window. Krsinich called this a *window splice* as opposed to the IDF *movement splice*. De Haan (2015, 27) suggested that perhaps the linking period should be in the middle of the old window which the Australian Bureau of Statistics (2016, 12) termed a *half splice*. Ivancic et al. (2010) suggested that the *average* of all links for the last period in the new window to the observations in the old window could be used as the linking factor. Diewert and Fox (2020) looked at these alternative methods for linking. *Average* or *mean linking* seems to be the safest strategy.

[17] For retrospective studies on upper level substitution bias for national CPIs, see Diewert et al. (2009a, 2009b), Huang et al. (2015) and Armknecht and Silver (2014). For studies of lower level substitution bias for a Lowe index, see Diewert et al. (2009a, 2009b) and Diewert (2014).

weight to the prices of products that may be unimportant in household budgets, which can lead to a biased Consumer Price Index.

There is a possible fifth method to avoid chain drift within a year when using a superlative index and that is to simply compute a sequence of 12 year over year monthly indexes so that say January prices in the previous year would be compared with January prices in the current year and so on. Handbury et al. (2013) used this methodological approach for the construction of year over year monthly superlative Japanese consumer price indexes using the Nikkei point of sale database. This database has monthly price and expenditure data covering the years 1988 to 2010 and contains 4.82 billion price and quantity observations. This type of index number was recommended in chapter 22 of the 2004 *Consumer Price Index Manual* as a valid year over year index that would avoid seasonality problems. However, central banks and other users require month-to-month CPIs in addition to year over year monthly CPIs and so the approach of Handbury, Watanabe and Weinstein does not solve the problems associated with the construction of superlative month-to-month indexes.

Many national statistical agencies are using web-scraping to collect large numbers of prices as a substitute for selective sampling of prices at the first stage of aggregation. Thus it is of interest to look at elementary indexes that depend only on prices, such as the Carli (1804), Dutot (1738) and Jevons (1865) indexes, and compare these indexes to superlative indexes; i.e., under what conditions will these indexes adequately approximate a superlative index.[18]

The two superlative indexes that we will consider in this chapter are the Fisher (1922) and the Törnqvist[19] indexes. The reasons for singling out these two indexes as preferred bilateral index number formulae are as follows: (i) both indexes can be given a strong justification from the viewpoint of the economic approach to index number theory; (ii) the Fisher index emerges as probably being the "best" index from the viewpoint of

[18] We will also look at the approximation properties of the CES price index with equal weights.

[19] The usual reference is Törnqvist (1936) but the index formula did not actually appear in this paper. It did appear explicitly in Törnqvist and Törnqvist (1937). It was listed as one of Fisher's (1922) many indexes: namely number 123. It was explicitly recommended as one of his top five ideal indexes by Warren Persons (1928, 86) so it probably should be called the *Persons index*. Theil (1967) developed a compelling descriptive statistics justification for the index. Superlative indexes are explained in Diewert (1976, 2021a).

the axiomatic or test approach to index number theory[20]; (iii) the Törnqvist index has a strong justification from the viewpoint of the stochastic approach to index number theory.[21] Thus there are strong cases for the use of these two indexes when making comparisons of prices between two periods when detailed price and quantity data are available.

When comparing two indexes, two methods for making the comparisons will be used: (i) use second-order Taylor series approximations to the index differences; (ii) the difference between two indexes can frequently be written as a covariance and it is possible in many cases to determine the likely sign of the covariance.[22]

When looking at scanner data from a retail outlet (or price and quantity data from a firm that uses dynamic pricing to price its products or services[23]), a fact emerges: if a product or a service is offered at a highly discounted price (i.e., it goes on sale), then the quantity sold of the product can increase by a very large amount. This empirical observation will allow us to make reasonable guesses about the signs of various covariances that express the difference between two indexes. If we are aggregating products that are close substitutes for each other, then a heavily discounted price may not only increase the *quantities sold* of the product but it may also increase the *expenditure share* of the sales in the list of products or services that are in scope for the index.[24] It turns out that the behavior of shares in response to discounted prices does make a difference in analyzing the differences between various indexes: in the context of highly substitutable products, a heavily discounted price will probably increase the market share of the product but if the products are weak substitutes (which is typically the case at higher levels of aggregation), then a discounted price will typically increase sales of the product but not increase its market share. These two cases (strong or weak substitutes) will play an important role in our analysis.

[20] See Diewert (1992).

[21] See Theil (1967, 136–137) or Chapter 4.

[22] This second method for making comparisons can be traced back to Bortkiewicz (1923).

[23] Airlines and hotels are increasingly using dynamic pricing; i.e., they change prices frequently.

[24] In the remainder of this chapter, we will speak of products but the same analysis applies to services.

Sections "Comparing CES Price Levels and Price Indexes" and "Using Means of Order r to Aggregate Price Ratios" look at relationships between the fixed base and chained Carli, Dutot, Jevons and CES (Constant Elasticity of Substitution) elementary indexes that do not use expenditure share or quantity information. These indexes are used by national statistical agencies at the first stage of aggregation when they calculate price indexes for components of their consumer price indexes in the case when quantity or value information is not available. It should be noted that we will start our analysis of various index number formulae by first developing the concept of a *price level*, which is an average of prices pertaining to a given period of time. A bilateral *price index* calculates *price change* between two periods. A price index could be a ratio of two price levels or it could be an average of price ratios, where the price of a good or service in the comparison period is in the numerator and the corresponding price in the base period is in the denominator. Comparing price levels for two periods is quite different from undertaking price comparisons over multiple periods. In the multiple period case, it turns out to be easier to compare price *levels* across periods rather than taking averages of price ratios as is done in the case of bilateral comparisons. Thus from the viewpoint of the economic approach to index number theory, it is simpler to target the estimation of unit cost functions rather than target the estimation of a ratio of unit cost functions. Once we have estimates for period by period price levels, we can easily form ratios of these estimates which will give us "normal" index numbers.

Section "Relationships Between Some Share Weighted Price Indexes" looks at the relationships between the Laspeyres, Paasche, Geometric Laspeyres, Geometric Paasche, Fisher and Törnqvist bilateral price indexes. Section "Relationships Between the Jevons, Geometric Laspeyres, Geometric Paasche and Törnqvist Price Indexes" investigates how close the unweighted Jevons index is to the Geometric Laspeyres P_{GL}^{t}, Geometric Paasche P_{GP}^{t} and Törnqvist P_{T}^{t} price indexes.

Section "Relationships Between Superlative Fixed Base Indexes and Geometric Indexes That Use Average Annual Shares as Weights" develops some relationships between the Törnqvist index and geometric indexes that use average *annual* shares as weights.

Section "To Chain or Not to Chain" looks at the differences between fixed base and chained Törnqvist indexes.

Multilateral indexes finally make their appearance in section "Relationships Between the Törnqvist Index and the GEKS and CCDI Multilateral Indexes": the fixed base Törnqvist index is compared to the GEKS

(Gini, Eltetö, Köves and Szulc) and GEKS-Törnqvist or CCDI (Caves, Christensen, Diewert and Inklaar) multilateral indexes.

Sections "Unit Value Price and Quantity Indexes" and "Quality Adjusted Unit Value Price and Quantity Indexes" compare Unit Value and Quality Adjusted Unit Value indexes to the Fisher index. It turns out that some multilateral indexes are actually quality adjusted unit value indexes as will be seen in section "Geary Khamis Multilateral Indexes". Section "Relationships Between Lowe and Fisher Indexes" compares the Lowe index to the Fisher index.

Section "Geary Khamis Multilateral Indexes" looks at the Geary Khamis multilateral index and shows that it is actually a special case of a quality adjusted unit value index.

Sections "Time Product Dummy Regressions: The Case of No Missing Observations" and "Time Product Dummy Regressions: The Case of Missing Observations" introduce Time Product Dummy multilateral indexes. Section "Time Product Dummy Regressions: The Case of No Missing Observations" assumes that there are no missing products in the window of time periods under consideration while section "Time Product Dummy Regressions: The Case of Missing Observations" deals with the case of missing products. Sections "Weighted Time Product Dummy Regressions: The Bilateral Case" and "Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations" introduce Weighted Time Product Dummy indexes for the case of two periods; the missing products case is considered in section "Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations". Finally, the Weighted Time Product Dummy multilateral indexes for T periods with missing products is discussed in section "Weighted Time Product Dummy Regressions: The General Case". Readers who are only interested in the general case can skip sections "Time Product Dummy Regressions: The Case of No Missing Observations"–"Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations" and just consider the general case in section "Weighted Time Product Dummy Regressions: The General Case".

Section "Linking Based on Relative Price Similarity" introduces a less familiar multilateral method that is based on linking observations that have the most similar structure of relative prices. This *similarity method* for linking observations has for the most part been used in the context of making cross-country comparisons. This class of methods depends on the choice of a measure of ***dissimilarity*** between the prices of two observations. The dissimilarity measure used in section "Linking Based on Relative Price Similarity" is Diewert's (2009) asymptotic linear measure of relative price dissimilarity.

A problem with the dissimilarity measure used in section "Linking Based on Relative Price Similarity" is that it requires positive prices for all products.[25] Thus in section "Inflation Adjusted Carry Forward and Backward Imputed Prices", a simple method for constructing imputed prices for missing products is described.

In section "Linking Based on Relative Price and Quantity Similarity", a new measure of relative price dissimilarity, the predicted share measure of relative price dissimilarity, is defined that does not require positive prices for all products in the two periods being compared. This new measure can be adapted to measures of dissimilarity between relative quantities. Section "Linking Based on Relative Price and Quantity Similarity" also introduces another method for constructing bilateral index number links between pairs of observations that have either proportional price vectors or proportional quantity vectors. This new method has some good axiomatic properties as will be seen in the following section "The Axiomatic Approach to Multilateral Price Levels".

Section "The Axiomatic Approach to Multilateral Price Levels" introduces an axiomatic or test approach to evaluate the properties of alternative multilateral methods for generating price and quantity levels cross multiple time periods. However, this section makes only a start on the axiomatic approach to evaluating alternative price levels for many time periods.

Section "Summary of Results" summarizes some of the more important results in this chapter.

The online Appendix [26] evaluates all of the above indexes for a grocery store scanner data set that is publicly available. This data set had a number of missing prices and quantities. Some of these missing prices may be due

[25] Products that are absent in both periods that are being compared can be ignored. However for products that are present in only one of the two comparison periods, the dissimilarity measure defined in section "Linking Based on Relative Price Similarity" requires that an imputed price for the missing products be constructed.

[26] The scanner data are available from Erwin Diewert (2021a, 2021b, 2021c), *Scanner Data Elementary Price Indexes and the Chain Drift Problem* (Discussion Paper 20-07). Vancouver School of Economics, The University of British Columbia.

to lack of sales or shortages of inventory. A general problem is how should the introduction of new products and the disappearance of (possibly) obsolete products be treated in the context of forming a consumer price index? Hicks (1940, 140) suggested a general approach to this measurement problem in the context of the economic approach to index number theory. His approach was to apply normal index number theory but estimate (or guess at) hypothetical prices that would induce utility maximizing purchasers of a related group of products to demand 0 units of unavailable products. With these virtual (or reservation or imputed) prices in hand, one can just apply normal index number theory using the augmented price data and the observed quantity data. The empirical example discussed in the online Appendix uses the scanner data that was used in Diewert and Feenstra (2017) for frozen juice products for a Dominick's store in Chicago for three years. This data set had 20 observations where $q_{tn} = 0$. For these 0 quantity observations, Diewert and Feenstra estimated positive Hicksian reservation prices for these missing price observations and these imputed prices are used in the empirical example in the Appendix. The Appendix lists the Dominick's data along with the estimated reservation prices. The Appendix also has tables and charts of the various index number formulae that are discussed in the main text of the study.

Comparing CES Price Levels and Price Indexes

In this section, we will begin our analysis by considering alternative methods by which the prices for N related products could be aggregated into an *aggregate price level* for the products for a given period.

We introduce some notation that will be used in the rest of the chapter. It is supposed that price and quantity data for N closely related products have been collected for T time periods.[27] Typically, a time period is a month. Denote the price of product n in period t as p_{tn} and the corresponding quantity during period t as q_{tn} for $n = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. Usually, p_{tn} will be the period t *unit value price* for product n in

[27] The T periods can be regarded as a window of observations, followed by another window of length T that has dropped the first period from the window and added the data of period $T + 1$ to the window. The literature on how to link the results of one window to the next window was briefly discussed in the introduction and is discussed at length in Diewert and Fox (2020).

period t; i.e., $p_{tn} = v_{tn}/q_{tn}$ where v_{tn} is the total value of product n that is sold or purchased during period t and q_{tn} is the total quantity of product n that is sold or purchased during period t. We assume that $q_{tn} \geq 0$ and $p_{tn} > 0$ for all t and n.[28] The restriction that all products have positive prices associated with them is a necessary one for much of our analysis since many popular index numbers are constructed using logarithms of prices and the logarithm of a zero price is not well defined. However, our analysis does allow for possible 0 quantities and values for some products for some time periods. Denote the period t strictly positive *price vector* as $p^t \equiv [p_{t1}, \ldots, p_{tN}] >> 0_N$ and nonnegative (and nonzero) *quantity vector* as $q^t \equiv [q_{t1}, ..., q_{tN}] > 0_N$ respectively, for $t = 1, \ldots, T$ where 0_N is an N dimensional vector of zeros. As usual, the inner product of the vectors p^t and q^t is denoted by $p^t \cdot q^t \equiv \sum_{n=1}^{N} p_{tn}q_{tn} > 0$. Define the period t sales (or expenditure) share for product n as $s_{tn} \equiv p_{tn}q_{tn}/p^t \cdot q^t$ for $n = 1, ..., N$ and $t = 1, ..., T$. The *period t sales or expenditure share vector* is defined as $s^t \equiv [s_{t1}, ..., s_{tN}] > 0_N$ for $t = 1, ..., T$.

In many applications, the N products will be closely related and they will have common units of measurement (by weight, or by volume or by "standard" package size). In this context, it is useful to define the period t "real" share for product n of total product sales or purchases, $S_{tn} \equiv q_{tn}/1_N \cdot q^t$ for $n = 1, ..., N$ and $t = 1, ..., T$ where 1_N is an N dimensional vector of ones. Denote the *period t real share vector* as $S^t \equiv [S_{t1}, ..., S_{tN}]$ for $t = 1, ..., T$.

Define a generic *product weighting vector* as $\alpha \equiv [\alpha_1, ..., \alpha_N]$. We assume that α has strictly positive components which sum to one; i.e., we assume that α satisfies:

$$\alpha \cdot 1_N = 1;\ \alpha >> 0_N. \tag{11.1}$$

Let $p \equiv [p_1, ..., p_N] >> 0_N$ be a strictly positive price vector. The corresponding *mean of order r of the prices p (with weights α)* or *CES price*

[28] In the case where $q_{tn} = 0$, then $v_{tn} = 0$ as well and hence $p_{tn} \equiv v_{tn}/q_{tn}$ is not well defined in this case. In the case where $q_{tn} = 0$, we will assume that p_{tn} is a positive imputed price. Imputed prices will be discussed in section "Inflation Adjusted Carry Forward and Backward Imputed Prices".

level, $m_{r,\alpha}(p)$ is defined as follows[29]:

$$m_{r,\alpha}(p) \equiv \left[\sum_{n=1}^{N} \alpha_n p_n^r\right]^{1/r}; \quad r \neq 0;$$
$$\equiv \prod_{n=1}^{N} (p_n)^{\alpha_n}; \quad r = 0. \tag{11.2}$$

It is useful to have a special notation for $m_{r,\alpha}(p)$ when $r = 1$:

$$p_\alpha \equiv \sum_{n=1}^{N} \alpha_n p_n = \alpha \cdot p. \tag{11.3}$$

Thus p_α is an α weighted arithmetic mean of the prices $p_1, p_2, ..., p_N$ and it can be interpreted as a *weighted Dutot price level*.[30]

From Schlömilch's (1858) Inequality,[31] we know that $m_{r,\alpha}(p) \geq m_{s,\alpha}(p)$ if $r \geq s$ and $m_{r,\alpha}(p) \leq m_{s,\alpha}(p)$ if $r \leq s$. However, we do not know how big the gaps are between these price levels for different r and s. When $r = 0$, $m_{0,\alpha}(p)$ becomes a weighted geometric mean or a *weighted* Jevons (1865) or *Cobb-Douglas price level* and it is of interest to know how much higher the weighted Dutot price level is than the corresponding weighted Jevons price level. Proposition 1 below provides an approximation to the gap between $m_{r,\alpha}(p)$ and $m_{1,\alpha}(p)$ for any r, including $r = 0$.

[29] Hardy et al. (1934, 12–13) refer to this family of means or averages as *elementary weighted mean values* and study their properties in great detail. The function $m_{r,\alpha}(p)$ can also be interpreted as a *Constant Elasticity of Substitution (CES) unit cost function* if $r \leq 1$. The corresponding utility or production function was introduced into the economics literature by Arrow et al. (1961). For additional material on CES functions, see Diewert (2021a), Feenstra (1994) and Diewert and Feenstra (2017).

[30] The ordinary Dutot (1738) price level for the period t prices p^t is defined as $p_D^t \equiv (1/N)\sum_{n=1}^{N} p_{tn}$. Thus it is equal to $m_{1,\alpha}(p^t)$ where $\alpha = (1/N)1_N$.

[31] See Hardy et al. (1934, 26) for a proof of this result.

Define the α *weighted variance* of $p/p_\alpha \equiv [p_1/p_\alpha, \ldots, p_N/p_\alpha]$ where p_α is defined by (11.3) as follows[32]:

$$\mathrm{var}_\alpha(p/p_\alpha) \equiv \sum_{n=1}^{N} \alpha_n[(p_n/p_\alpha) - 1]^2. \tag{11.4}$$

Proposition 1: Let $p >> 0_N$, $\alpha >> 0_N$ and $\alpha \cdot 1_N = 1$. Then $m_{r,\alpha}(p)/m_{1,\alpha}(p)$ is approximately equal to the following expression for any r.

$$m_{r,\alpha}(p)/m_{1,\alpha}(p) \approx 1 + (1/2)(r-1)\mathrm{var}_\alpha(p/p_\alpha) \tag{11.5}$$

where $\mathrm{var}_\alpha(p/p_\alpha)$ is defined by (11.4). The expression on the right hand side of (11.5) uses a second-order Taylor series approximation to $m_{r,\alpha}(p)$ around the equal price point $p = p_\alpha 1_N$ where p_α is defined by (11.3).[33]

Proof: Straightforward calculations show that the level, vector of first-order partial derivatives and matrix of second-order partial derivatives of $m_{r,\alpha}(p)$ evaluated at the equal price point $p = p_\alpha 1_N$ are equal to the following expressions: $m_{r,\alpha}(p_\alpha 1_N) = p_\alpha \equiv \alpha \cdot p$; $\nabla_p m_{r,\alpha}(p_\alpha 1_N) = \alpha$; $\nabla^2_{pp} m_{r,\alpha}(p_\alpha 1_N) = (p_\alpha)^{-1}(r-1)(\hat{\alpha} - \alpha\alpha^T)$ where $\hat{\alpha}$ is a diagonal N by N matrix with the elements of the column vector α running down the main diagonal and α^T is the transpose of the column vector α. Thus $\alpha\alpha^T$ is a rank one N by N matrix.

Thus the second-order Taylor series approximation to $m_{r,\alpha}(p)$ around the point $p = p_\alpha 1_N$ is given by the following expression:

$$\begin{aligned} m_{r,\alpha}(p) &\approx p_\alpha + \alpha \cdot (p - p_\alpha 1_N) \\ &\quad + \frac{1}{2}(p - p_\alpha 1_N)^T (p_\alpha)^{-1}(r-1)\left(\hat{\alpha} - \alpha\alpha^T\right)(p - p_\alpha 1_N) \\ &= p_\alpha + \frac{1}{2}(p_\alpha)^{-1}(r-1)(p - p_\alpha 1_N)^T (p_\alpha)^{-1}\left(\hat{\alpha} - \alpha\alpha^T\right)(p - p_\alpha 1_N) \end{aligned} \tag{11.6}$$

[32] Note that the α weighted mean of p/p_α is equal to $\sum_{n=1}^{N} \alpha_n p_n/p_\alpha = 1$. Thus (11.4) defines the corresponding weighted variance.

[33] For alternative approximations for the differences between mean of order r averages, see Vartia (1978, 278–279). Vartia's approximations involve variances of logarithms of prices, whereas our approximations involve variances of deflated prices. Our analysis is a variation on his pioneering analysis.

using (11.1) and (11.3)

$$= p_\alpha\left[1 + \frac{1}{2}(r-1)(p_\alpha)^{-2}(p - p_\alpha 1_N)^T\left(\hat{\alpha} - \alpha\alpha^T\right)(p - p_\alpha 1_N)\right]$$
$$= m_{1,\alpha}(p)\left[1 + \frac{1}{2}(r-1)\ Var_\alpha\left(\frac{p}{p_\alpha}\right)\right]$$

using (11.2), (11.3) and (11.4).

Q.E.D.

The approximation (11.6) also holds if $r = 0$. In this case (11.6) becomes the following approximation[34]:

$$\begin{aligned} m_{0,\alpha}(p) &\equiv \prod_{n=1}^{N}(p_n)^{\alpha_n} \\ &\approx m_{1,\alpha}(p)\left[1 - \frac{1}{2}\text{var}_\alpha\left(\frac{p}{p_\alpha}\right)\right] \\ &= m_{1,\alpha}(p)\left\{1 - \frac{1}{2}\sum_{n=1}^{N}\alpha_n\left[\frac{p_n}{p_\alpha} - 1\right]^2\right\} \end{aligned} \tag{11.7}$$

using (11.4)

$$= \left[\sum_{n=1}^{N}\alpha_n p_n\right]\left\{1 - \frac{1}{2}\sum_{n=1}^{N}\alpha_n\left[\frac{p_n}{p_\alpha} - 1\right]^2\right\}$$

using (11.2) for $r = 1$

$$\leq \sum_{n=1}^{N}\alpha_n p_n.$$

Thus the bigger is the variation in the N prices $p_1, \ldots, p_N$, the bigger will be $\text{var}_\alpha(p/p_\alpha)$ and the more the weighted arithmetic mean of

[34] Note that $m_{0,\alpha}(p)$ can be regarded as a weighted Jevons (1865) price level or a Cobb Douglas (1928) price level. Similarly, $p_\alpha \equiv m_{1,\alpha}(p)$ can be regarded as a weighted Dutot (1738) price level or a Leontief (1936) price level.

the prices, $\sum_{n=1}^{N} \alpha_n p_n$, will be greater than the corresponding weighted geometric mean of the prices, $\prod_{n=1}^{N} (p_n)^{\alpha_n}$. Note that if all of the p_n are equal, then $\mathrm{var}_\alpha(p/p_\alpha)$ will be equal to 0 and the approximations in (11.6) and (11.7) become exact equalities.

At this point, it is useful to define the Jevons (1865) and Dutot (1738) period t *price levels* for the prices in our window of observations, p_J^t and p_D^t, and the corresponding Jevons and Dutot *price indexes*, P_J^t and P_D^t, for $t = 1, \ldots, T$:

$$p_D^t \equiv \sum_{n=1}^{N} \frac{1}{N} p_{tn}; \tag{11.8}$$

$$p_J^t \equiv \prod_{n=1}^{N} p_{tn}^{1/N}; \tag{11.9}$$

$$P_D^t \equiv \frac{p_D^t}{p_D^1}; \tag{11.10}$$

$$P_J^t \equiv \frac{p_J^t}{p_J^1} = \prod_{n=1}^{N} \left(\frac{p_{tn}}{p_{1n}}\right)^{1/N}. \tag{11.11}$$

Thus the period t price index is simply the period t price level divided by the corresponding period 1 price level. Note that the Jevons price index can also be written as the geometric mean of the long-term price ratios (p_{tn}/p_{1n}) between the period t prices relative to the corresponding period 1 prices.

The *weighted Dutot and Jevons period t price levels* using a weight vector α which satisfies the restrictions (11.1), $p_{D\alpha}^t$ and $p_{J\alpha}^t$, are defined by (11.12) and (11.13) and the corresponding *weighted Dutot and Jevons period t price indexes*, $P_{D\alpha}^t$[35] and $P_{J\alpha}^t$,[36] are defined by (11.14) and (11.15) for $t = 1, \ldots, T$:

$$p_{D\alpha}^t \equiv \sum_{n=1}^{N} \alpha_n p_{tn} = m_{1,\alpha}(p^t); \tag{11.12}$$

[35] A weighted Dutot index can also be interpreted as a Lowe (1823) index.

[36] This type of index is frequently called a *Geometric Young index*; see Armknecht and Silver (2014, 4–5).

$$p^t_{J\alpha} \equiv \prod_{n=1}^{N} (p_{tn})^{\alpha_n} = m_{0,\alpha}(p^t); \tag{11.13}$$

$$P^t_{D\alpha} \equiv \frac{p^t_{D\alpha}}{p^1_{D\alpha}} = \frac{\alpha \cdot p^t}{\alpha \cdot p^1}; \tag{11.14}$$

$$P^t_{J\alpha} \equiv \frac{p^t_{J\alpha}}{p^1_{J\alpha}} = \prod_{n=1}^{N} \left(\frac{p_{tn}}{p_{1n}}\right)^{\alpha_n}. \tag{11.15}$$

Obviously (11.12)–(11.15) reduce to definitions (11.8)–(11.11) if $\alpha = (1/N)1_N$. We can use the approximation (11.7) for $p = p^1$ and $p = p^t$ in order to obtain the following approximate relationship between the weighted Dutot price index for period t, $P^t_{D\alpha}$, and the corresponding weighted Jevons index, $P^t_{J\alpha}$.

$$\begin{aligned} P^t_{J\alpha} &\equiv \frac{p^t_{J\alpha}}{p^1_{J\alpha}}; \quad t = 1, \ldots, T \\ &= \frac{m_{0,\alpha}(p^t)}{m_{0,\alpha}(p^1)} \end{aligned} \tag{11.16}$$

using (11.2) and (11.13)

$$\approx \frac{m_{1,\alpha}(p^t)\left\{1 - \frac{1}{2}\sum_{n=1}^{N}\alpha_n\left[\frac{p_{tn}}{p^t_\alpha} - 1\right]^2\right\}}{m_{1,\alpha}(p^1)\left\{1 - \frac{1}{2}\sum_{n=1}^{N}\alpha_n\left[\frac{p_{1n}}{p^1_\alpha} - 1\right]^2\right\}}$$

using (11.7) for $p = p^t$ and $p = p^1$ where $p^t_\alpha \equiv \alpha \cdot p^t$ and $p^1_\alpha \equiv \alpha \cdot p^1$

$$\begin{aligned} &= \frac{P^t_{D\alpha}\left\{1 - \frac{1}{2}\sum_{n=1}^{N}\alpha_n\left[\frac{p_{tn}}{p^t_\alpha} - 1\right]^2\right\}}{\left\{1 - \frac{1}{2}\sum_{n=1}^{N}\alpha_n\left[\frac{p_{1n}}{p^1_\alpha} - 1\right]^2\right\}} \\ &= \frac{P^t_{D\alpha}\left[1 - \frac{1}{2}\mathrm{var}_\alpha\left(\frac{p^t}{p^t_\alpha}\right)\right]}{\left[1 - \frac{1}{2}\mathrm{var}_\alpha\left(\frac{p^1}{p^1_\alpha}\right)\right]}. \end{aligned}$$

In the elementary index context where there are no trends in prices in diverging directions, it is likely that $\mathrm{var}_\alpha(p^t / p_\alpha^t)$ will be approximately equal to $\mathrm{var}_\alpha(p^1 / p_\alpha^1)$.[37] Under this condition, *the weighted Jevons price index* $P_{J\alpha}^t$ *is likely to be approximately equal to the corresponding weighted Dutot price index*, $P_{D\alpha}^t$. Of course, this approximate equality result extends to the case where $\alpha = (1/N)1_N$ and so it is likely that the Dutot price indexes P_D^t are approximately equal to their Jevons price index counterparts, P_J^t.[38] However, if the variance of the deflated period 1 prices is unusually large (small), then there will be a tendency for P_J^t to exceed (to be less than) P_D^t for $t > 1$.

At higher levels of aggregation where the products may not be very similar,[39] it is likely that there will be *divergent trends in prices* over time. In this case, we can expect $\mathrm{var}_\alpha(p^t / p_\alpha^t)$ to exceed $\mathrm{var}_\alpha(p^1 / p_\alpha^1)$. Thus using (11.16) under these circumstances leads to the likelihood that the weighted index $P_{J\alpha}^t$ will be significantly lower than $P_{D\alpha}^t$. Similarly, under the *diverging trends in prices hypothesis*, we can expect the ordinary Jevons index P_J^t to be lower than the ordinary Dutot index P_D^t.[40]

We conclude this section by finding an approximate relationship between a CES price index and the corresponding weighted Dutot price index $P_{D\alpha}^t$. This approximation result assumes that econometric estimates for the parameters of the CES unit cost function $m_{r,\alpha}(p)$ defined by (11.2) are available so that we have estimates for the weighting vector α (which we assume satisfies the restrictions [11.1]) and the parameter r which we assume satisfies $r \le 1$.[41] The *CES period t price levels* using a weight vector α which satisfies the restrictions (11.1) and an $r \le 1$,

[37] Note that the vectors p^t / p_α^t and p^1 / p_α^1 are price vectors that are divided by their α weighted arithmetic means. Thus these vectors have eliminated general inflation between periods 1 and t.

[38] The same approximate inequalities hold for the weighted case. An approximation result similar to (11.16) for the equal weights case where $\alpha = (1/N)1_N$ was first obtained by Carruthers et al. (1980, 25). See Diewert (2021b), Eq. (11.16).

[39] If the products are not very similar, then the Dutot index should not be used since it is not invariant to changes in the units of measurement.

[40] Furthermore, as we shall see later, the Dutot index can be viewed as a fixed basket index where the basket is a vector of ones. Thus it is subject to substitution bias that will show up under the divergent price trends hypothesis.

[41] These restrictions imply that $m_{r,\alpha}(p)$ is a linearly homogeneous, nondecreasing and concave function of the price vector p. These restrictions must be satisfied if we apply the economic approach to price index theory.

$p^t_{\text{CES}\alpha,r}$, and the corresponding *CES period t price indexes*, $P^t_{\text{CES}\alpha,r}$, are defined as follows for $t = 1, \ldots, T$:

$$p^t_{\text{CES}\alpha,r} \equiv \left[\sum_{n=1}^{N} \alpha_n p^r_{tn}\right]^{1/r} = m_{r,\alpha}\left(p^t\right); \tag{11.17}$$

$$P^t_{\text{CES}\alpha,r} \equiv \frac{p^t_{\text{CES}\alpha,r}}{p^1_{\text{CES}\alpha,r}} = \frac{m_{r,\alpha}\left(p^t\right)}{m_{r,\alpha}\left(p^1\right)}. \tag{11.18}$$

Now use the approximation (11.6) for $p = p^1$ and $p = p^t$ in order to obtain the following approximate relationship between the weighted Dutot price index for period t, $P^t_{D\alpha}$, and the corresponding period t CES index, $P^t_{\text{CES}\alpha,r}$ for $t = 1, \ldots, T$:

$$\begin{aligned} P^t_{\text{CES}\alpha,r} &\equiv \frac{p^t_{\text{CES}\alpha,r}}{p^1_{\text{CES}\alpha,r}}, \\ &= \frac{m_{r,\alpha}\left(p^t\right)}{m_{r,\alpha}\left(p^1\right)} \end{aligned} \tag{11.19}$$

using (11.18)

$$\approx \frac{m_{1,\alpha}\left(p^t\right)}{m_{1,\alpha}\left(p^1\right)} \frac{\left[1 + \frac{1}{2}(r-1)\text{var}_\alpha \frac{p^t}{p^t_\alpha}\right]}{\left[1 + \frac{1}{2}(r-1)\text{var}_\alpha \frac{p^1}{p^1_\alpha}\right]}$$

$$= \frac{P^t_{D\alpha}\left\{1 + \frac{1}{2}(r-1)\sum_{n=1}^{N} \alpha_n\left[\left(\frac{p_{tn}}{p^t_\alpha}\right) - 1\right]^2\right\}}{\left\{1 + \frac{1}{2}(r-1)\sum_{n=1}^{N} \alpha_n\left[\left(\frac{p_{1n}}{p^1_\alpha}\right) - 1\right]^2\right\}}$$

where we used definitions (11.4), (11.12) and (11.14) to establish the last equality in (11.19). Again, in the elementary index context with no diverging trends in prices, we could expect $\text{var}_\alpha\left(p^t/p^t_\alpha\right) \approx \text{var}_\alpha\left(p^1/p^1_\alpha\right)$ for $t = 2, \ldots, T$. Using this assumption about the approximate constancy of the (weighted) variance of the deflated prices over time, and using (11.16) and (11.19), we obtain the following approximations for $t = 2, 3, \ldots, T$:

$$P^t_{\text{CES}\alpha,r} \approx P^t_{J\alpha} \approx P^t_{D\alpha}. \tag{11.20}$$

Thus, under the assumption of approximately *constant variances* for deflated prices, the CES, weighted Jevons and weighted Dutot price indexes should approximate each other fairly closely, provided that the same weighting vector α is used in the construction of these indexes.[42]

The parameter r which appears in the definition of the CES unit cost function is related to the *elasticity of substitution* σ; i.e., it turns out that $\sigma = 1 - r$.[43] Thus as r takes on values from 1 to $-\infty$, σ will take on values from 0 to $+\infty$. In the case where the products are closely related, typical estimates for σ range from 1 to 10. If we substitute $\sigma = 1 - r$ into the approximations (11.19), we obtain the following approximations for $t = 1, \ldots, T$:

$$P^t_{\text{CES}\alpha,r} \approx \frac{P^t_{D\alpha}\left[1 - \frac{1}{2}\sigma \text{var}_\alpha\left(\frac{p^t}{p^t_\alpha}\right)\right]}{\left[1 - \frac{1}{2}\sigma \text{var}_\alpha\left(\frac{p^1}{p^1_\alpha}\right)\right]}. \tag{11.21}$$

The approximations in (11.21) break down for large and positive σ (or equivalently, for very negative r); i.e., the expressions in square brackets on the right hand sides of (11.21) will pass through 0 and become meaningless as σ becomes very large. These approximations become increasingly accurate as σ approaches 0 (or as r approaches 1). Of course, the approximations also become more accurate as the dispersion of prices within a period becomes smaller. For σ between 0 and 1 and with "normal" dispersion of prices, the approximations in (11.21) should be reasonably good. However, as σ becomes larger, the expressions in square brackets will become closer to 0 and the approximations in (11.21) will become more volatile and less accurate as σ increases from an initial 0 value.

If the products in the aggregate are not very similar, it is more likely that there will be *divergent trends in prices* over time and in this case, we can expect $\text{var}_\alpha(p^t/p^t_\alpha)$ to exceed $\text{var}_\alpha(p^1/p^1_\alpha)$. In this case, the approximate *equalities* (11.20) will no longer hold. In the case where the

[42] Again, the approximate relationship $P^t_{\text{CES}\alpha,r} \approx P^t_{D\alpha}$ may not hold if the variance of the prices in the base period, $\text{var}_\alpha\left(p^1/p^1_\alpha\right)$, is unusually large or small. Also, under the diverging trends in prices assumption, $\text{var}_\alpha(p^t/p^t_\alpha)$ will tend to increase relative to $\text{var}_\alpha\left(p^1/p^1_\alpha\right)$ and the approximate equalities in (11.20) will become inequalities.

[43] See Feenstra (1994, 158) or Eq. (11.115) in Diewert (2021a).

elasticity of substitution σ is greater than 1 (so $r < 0$) and $\mathrm{var}_\alpha\left(p^t / p^t_\alpha\right) > \mathrm{var}_\alpha\left(p^1 / p^1_\alpha\right)$, we can expect that $P^t_{\mathrm{CES}\alpha,r} < P^t_{D\alpha}$ and the gaps between these two indexes will grow bigger over time as $\mathrm{var}_\alpha\left(p^t / p^t_\alpha\right)$ grows larger than $\mathrm{var}_\alpha\left(p^1 / p^1_\alpha\right)$.

In the following section, we will use the mean of order r function to aggregate the price ratios p_{tn} / p_{1n} into an aggregate price index for period t directly; i.e., we will not construct *price levels* as a preliminary step in the construction of a *price index*.

Using Means of Order r to Aggregate Price Ratios

In the previous section, we compared various elementary indexes using approximate relationships between price levels constructed by using means of order r to construct the aggregate *price levels*. In this section, we will develop approximate relationships between price indexes constructed by using means of order r to aggregate over *price ratios*.

In what follows, it is assumed that the weight vector α satisfies conditions (1); i.e., $\alpha >> 0_N$ and $\alpha 1_N = 1$. Define the *mean of order r price index for period t* (relative to period 1), $P^t_{r,\alpha}$, as follows for $t = 1, \ldots, T$:

$$P^t_{r,\alpha} \equiv \left[\sum_{n=1}^{N} \alpha_N \left(\frac{p_{tn}}{p_{1n}}\right)^r\right]^{1/r}; \quad r \neq 0;$$

$$\equiv \prod_{n=1}^{N} \left(\frac{p_{tn}}{p_{1n}}\right)^{\alpha_n}; \quad r = 0. \tag{11.22}$$

When $r = 1$ and $\alpha = (1/N)1_N$, then $P^t_{r,\alpha}$ becomes the *fixed base* Carli (1804) *price index* (for period t relative to period 1), P^t_C, defined as follows for $t = 1, \ldots, T$:

$$P^t_C \equiv \sum_{n=1}^{N} \frac{1}{N} \left(\frac{p_{tn}}{p_{1n}}\right). \tag{11.23}$$

With a general α and $r = 1$, $P^t_{r,\alpha}$ becomes the *fixed base weighted Carli price index*, $P^t_{C\alpha}$,[44] defined as follows for $t = 1, \ldots, T$:

$$P^t_{C\alpha} \equiv \sum_{n=1}^{N} \alpha_n \left(\frac{p_{tn}}{p_{1n}} \right). \tag{11.24}$$

Using (11.24), it can be seen that the α weighted mean of the period t long-term price ratios p_{tn}/p_{1n} divided by $P^t_{C\alpha}$ is equal to 1; i.e., we have for $t = 1, \ldots, T$:

$$\sum_{n=1}^{N} \alpha_n \left(\frac{p_{tn}}{p_{1n} P^t_{C\alpha}} \right) = 1. \tag{11.25}$$

Denote the α weighted variance of the deflated period t price ratios $p_{tn}/p_{1n} P^t_{C\alpha}$ as $\mathrm{var}_\alpha\left(p^t/p^1 P^t_{C\alpha}\right)$ and define it as follows for $t = 1, \ldots, T$:

$$\mathrm{var}_\alpha \left(\frac{p^t}{p^1 P^t_{C\alpha}} \right) \equiv \sum_{n=1}^{N} \alpha_n \left[\left(\frac{p_{tn}}{p_{1n} P^t_{C\alpha}} \right) - 1 \right]^2. \tag{11.26}$$

Proposition 2: Let $p >> 0_N$, $\alpha >> 0_N$ and $\alpha 1_N = 1$. Then $P^t_{r,\alpha}/P^t_{1,\alpha} = P^t_{r,\alpha}/P^t_{C\alpha}$ is approximately equal to the following expression for any r for $t = 1, \ldots, T$:

$$\frac{P^t_{r,\alpha}}{P^t_{C\alpha}} \approx 1 + \frac{1}{2}(r-1)\mathrm{var}_\alpha \left(\frac{p^t}{p^1 P^t_{C\alpha}} \right) \tag{11.27}$$

where $P^t_{r,\alpha}$ is the mean of order r price index (with weights α) defined by (11.22), $P^t_{C\alpha}$ is the α weighted Carli index defined by (11.24) and $\mathrm{var}_\alpha\left(p^t/p^1 P^t_{C\alpha}\right)$ is the α weighted variance of the deflated long-term price ratios $\left(p_{tn}/p_{1n}\right) P^t_{C\alpha}$ defined by (11.26).

Proof: Replace the vector p in Proposition 1 by the vector $\left[p_{t1}/p_{11}, p_{t2}/p_{12}, \ldots p_{tN}/p_{1N}\right]$.[45] Then the ratio $m_{r,\alpha}(p)/m_{1,\alpha}(p)$

[44] This type of index is due to Arthur Young (1812, 72) and so we could call this index the *Young index*, $P^t_{Y\alpha}$.

[45] In Proposition 1, some prices in either period could be 0. However, Proposition 2 requires that all period 1 prices be positive.

which appears on the left hand side of (11.5) becomes the ratio $P^t_{r,\alpha} \big/ P^t_{1,\alpha} = P^t_{r,\alpha} / P^t_{C\alpha}$ using definitions (11.22) and (11.24). The terms p_α and $\text{var}_\alpha(p / p_\alpha)$ which appear on the right hand side of (11.5) become $P^t_{C\alpha}$ and $\text{var}_\alpha(p^t / p^1 P^t_{C\alpha})$ respectively. With these substitutions, (11.5) becomes (11.27) and we have established Proposition 2.

Q.E.D.

It is useful to look at the special case of (11.27) when $r = 0$. In this case, using definitions (11.22) and (11.15), we can establish the following equalities for $t = 1, \ldots, T$:

$$P^t_{0,\alpha} \equiv \prod_{n=1}^{N} \left(\frac{p_{tn}}{p_{1n}} \right)^{\alpha_n} = P^t_{J\alpha} \tag{11.28}$$

where $P^t_{J\alpha}$ is the period t *weighted Jevons* or *Cobb Douglas price index* defined by (11.15) in the previous section.[46] Thus when $r = 0$, the approximations defined by (11.27) become the following approximations for $t = 1, \ldots, T$:

$$\frac{P^t_{J\alpha}}{P^t_{C\alpha}} \approx 1 - \frac{1}{2}\text{var}_\alpha \left(\frac{p^t}{p^1 P^t_{C\alpha}} \right). \tag{11.29}$$

Thus the bigger is the α weighted variance of the deflated period t long-term price ratios, $(p_{t1} / p_{11}) / P^t_{C\alpha}, \ \ldots, (p_{tN} / p_{1N}) / P^t_{C\alpha}$, the more the period t weighted Carli index $P^t_{C\alpha}$ will exceed the corresponding period t weighted Jevons index $P^t_{J\alpha}$.

When $\alpha = (1 / N)1_N$, the approximations (11.29) become the following approximate relationships between the period t *Carli index* P^t_C defined by (11.23) and the period t *Jevons index* P^t_J defined by (11.11)

[46] Again, recall that Armknecht and Silver (2014, 4) call this index the Geometric Young index.

for $t = 1, \ldots, T$[47]:

$$\frac{P_J^t}{P_C^t} \approx 1 - \frac{1}{2}\text{var}_{(1/N)1}\left(\frac{p^t}{p^1 P_C^t}\right)$$
$$= 1 - \frac{1}{2}\sum_{n=1}^{N}\frac{1}{N}\left[\frac{p_{tn}}{p_{1n}P_C^t} - 1\right]^2. \quad (11.30)$$

Thus the Carli price indexes P_C^t will exceed their Jevons counterparts P_J^t (unless $p^t = \lambda_t p^1$ in which case prices in period t are proportional to prices in period 1 and in this case, $P_C^t = P_J^t$).[48] This is an important result, since from an axiomatic perspective, the Jevons price index has much better properties than the corresponding Carli indexes[49] and in particular, typically *chaining Carli indexes will lead to large upward biases as compared to their Jevons counterparts*.

The results in this section can be summarized as follows: holding the weight vector α constant, the weighted Jevons price index for period t, $P_{J\alpha}^t$ will lie below the corresponding weighted Carli index, $P_{C\alpha}^t$ (unless all prices move in a proportional manner, in which case $P_{J\alpha}^t$ will equal $P_{C\alpha}^t$) with the gap growing as the α weighted variance of the deflated price ratios, $(p_{t1}/p_{11})/P_{C\alpha}^t, \ldots, (p_{tN}/p_{1N})/P_{C\alpha}^t$, increases.[50]

[47] Results that are essentially equivalent to (11.30) were first obtained by Dalén (1992) and Diewert (1995). The approximations in (11.27) and (11.29) for weighted indexes are new. Vartia and Suoperä (2018, 5) derived alternative approximations. The analysis in this section is similar to Vartia's (1978, 276–289) analysis of Fisher's (1922) five-tined fork.

[48] From Schlömilch's Inequality, we know that P_C is always equal to or greater than P_J; the approximate result (11.30) provides an indication of the size of the gap between the two indexes.

[49] See Diewert (1995, 2021b) and Reinsdorf (2007) on the axiomatic approach to equally weighted elementary indexes. The Jevons index emerges as the best index from the viewpoint of the axiomatic approach.

[50] Since the Jevons price index has the best axiomatic properties, this result implies that CPI compilers should avoid the use of the Carli index in the construction of a CPI. This advice goes back to Fisher (1922, 29–30). Since the Dutot index will approximate the corresponding Jevons index provided that the products are similar and there are no systematic divergent trends in prices, Dutot indexes can be satisfactory at the elementary level. If the products are not closely related, Dutot indexes become problematic since they are not invariant to changes in the units of measurement. Moreover, in the case of

In the following section, we turn our attention to weighted price indexes where the weights are not exogenous constants but depend on observed sales or expenditure shares.

Relationships Between Some Share Weighted Price Indexes

In this section (and in subsequent sections), we will look at comparisons between price indexes that use information on the observed expenditure or sales shares of products in addition to price information. Recall that $s_{tn} \equiv p_{tn}q_{tn}/p^t \cdot q^t$ for $n = 1, \ldots, N$ and $t = 1, \ldots, T$.

The *fixed base* Laspeyres (1871) *price index* for period t, P_L^t, is defined as the following base period share weighted *arithmetic* average of the price ratios, p_{tn}/p_{1n}, for $t = 1, \ldots, T$:

$$P_L^t \equiv \sum_{n=1}^{N} s_{1n}\left(\frac{p_{tn}}{p_{1n}}\right). \qquad (11.31)$$

It can be seen that P_L^t is a weighted Carli index $P_{C\alpha}^t$ of the type defined by (11.24) in the previous section where $\alpha \equiv s^1 \equiv [s_{11}, s_{12}, \ldots, s_{1N}]$. We will compare P_L^t with its weighted geometric mean counterpart, P_{GL}^t, which is a weighted Jevons index $P_{J\alpha}^t$ where the weight vector is $\alpha = s^1$. Thus the logarithm of the *fixed base Geometric Laspeyres price index* is defined as follows for $t = 1, \ldots, T$[51]:

$$\ln P_{\text{GL}}^t \equiv \sum_{n=1}^{N} s_{1n} \ln\left(\frac{p_{tn}}{p_{1n}}\right). \qquad (11.32)$$

Since P_{GL}^t and P_L^t are weighted geometric and arithmetic means of the price ratios p_{tn}/p_{1n} (using the weights in the period 1 share vector s^1), Schlömilch's inequality implies that $P_{\text{GL}}^t \le P_L^t$ for $t = 1, \ldots, T$. The inequalities (11.29), with $\alpha = s^1$, give us approximations to the gaps between the $P_{\text{GL}}^t = P_{J\alpha}^t$ and the $P_{C\alpha}^t = P_L^t$. Thus we have the following

nonsimilar products, divergent trends in prices become more probable and, using (11.16), the Dutot index will tend to be above the corresponding Jevons index.

[51] Vartia (1978, 272) used the terms "geometric Laspeyres" and "geometric Paasche" to describe the indexes defined by (11.32) and (11.35).

approximate equalities for $\alpha = s^1$ and $t = 1, \ldots, T$:

$$\frac{P^t_{\text{GL}}}{P^t_L} \approx 1 - \frac{1}{2}\text{var}_\alpha\left(\frac{p^t}{p^1 P^t_L}\right) = 1 - \frac{1}{2}\sum_{n=1}^{N} s_{1n}\left[\left(\frac{p_{tn}}{p_{1n}P^t_L}\right) - 1\right]^2. \quad (11.33)$$

The *fixed base* Paasche (1874) *price index* for period t, P^t_P, is defined as the following period t share weighted *harmonic* average of the price ratios, p_{tn}/p_{1n}, for $t = 1, \ldots, T$:

$$P^t_P \equiv \left[\sum_{n=1}^{N} s_{tn}\left(\frac{p_{tn}}{p_{1n}}\right)^{-1}\right]^{-1}. \quad (11.34)$$

We will compare P^t_P with its weighted geometric mean counterpart, P^t_{GP}, which is a weighted Jevons index $P^t_{J\alpha}$ where the weight vector is $\alpha = s^t$. The logarithm of the *fixed base Geometric Paasche price index* is defined as follows for $t = 1, \ldots, T$:

$$\ln P^t_{\text{GP}} \equiv \sum_{n=1}^{N} s_{tn}\ln\left(\frac{p_{tn}}{p_{1n}}\right). \quad (11.35)$$

Since P^t_{GP} and P^t_P are weighted geometric and harmonic means of the price ratios p_{tn}/p_{1n} (using the weights in the period t share vector s^t), Schlömilch's inequality implies that $P^t_P \le P^t_{\text{GP}}$ for $t = 1, \ldots, T$. However, we cannot apply the inequalities (11.29) directly to give us an approximation to the *size* of the gap between P^t_{GP} and P^t_P. Viewing definition (11.34), it can be seen that the reciprocal of P^t_P is a period t share weighted average of the reciprocals of the long-term price ratios, $p_{11}/p_{t1}, p_{12}/p_{t2}, \ldots, p_{1N}/p_{tN}$. Thus using definition (11.34), we have the following equations and inequalities for $\alpha = s^t$ and $t = 1, \ldots, T$:

$$[P^t_P]^{-1} = \sum_{n=1}^{N} s_{tn}\left(\frac{p_{1n}}{p_{tn}}\right) \ge \prod_{n=1}^{N}\left(\frac{p_{1n}}{p_{tn}}\right)^{s_{tn}}$$

$$= \left(P_{GP}^t\right)^{-1} \tag{11.36}$$

using definitions (11.35)

where the inequalities in (11.36) follow from Schlömilch's inequality; i.e., a weighted arithmetic mean is always equal to or greater than the corresponding weighted geometric mean. Note that the first equation in (11.36) implies that the period t share weighted mean of the reciprocal price ratios, p_{1n}/p_{tn}, is equal to the reciprocal of P_P^t. Now adapt the approximate equalities (11.29) in order to establish the following approximate equalities for $t = 1, \ldots, T$:

$$\frac{\left(P_{GP}^t\right)^{-1}}{\left(P_P^t\right)^{-1}} \approx 1 - \frac{1}{2}\sum_{n=1}^{N} s_{tn}\left[\left(\frac{p_{1n}}{p_{tn}\left[P_P^t\right]^{-1}}\right) - 1\right]^2 \tag{11.37}$$

The approximate equalities (11.37) may be rewritten as follows for $t = 1, \ldots, T$:

$$P_{GP}^t \approx \frac{P_P^t}{1 - \frac{1}{2}\sum_{n=1}^{N} s_{tn}\left[\left(\frac{p_{1n}P_P^t}{p_{tn}}\right) - 1\right]^2}. \tag{11.38}$$

Thus for $t = 1, \ldots, T$, we have $P_{GP}^t \geq P_P^t$ (and the approximate equalities (11.38) measure the gaps between these indexes) and $P_{GP}^t \leq P_L^t$(and the approximate equalities (11.33) measure the gaps between these indexes). Later we will show that the inequalities $P_{GP}^t \leq P_{GL}^t$ are likely if the N products are close substitutes for each other.

Suppose that prices in period t are proportional to the corresponding prices in period 1 so that $p^t = \lambda_t p^1$ where λ_t is a positive scalar. Then it is straightforward to show that $P_P^t = P_{GP}^t = P_{GL}^t = P_L^t = \lambda_t$ and the implicit error terms for equation t in (11.33) and (11.38) are equal to 0.

Define the period t fixed base Fisher (1922) and Törnqvist Theil price indexes, P_F^t and P_T^t, as the following geometric means for $t = 1, \ldots, T$:

$$P_F^t \equiv \left[P_L^t P_P^t\right]^{1/2}, \tag{11.39}$$

$$P_T^t \equiv \left[P_{GL}^t P_{GP}^t\right]^{1/2}. \tag{11.40}$$

Thus P_F^t is the geometric mean of the period t fixed base Laspeyres and Paasche price indexes while P_T^t is the geometric mean of the period

t fixed base geometric Laspeyres and geometric Paasche price indexes. Now use the approximate equalities in (11.33) and (11.38) and substitute these equalities into (11.40) in order to obtain the following approximate equalities between P_T^t and P_F^t for $t = 1, \ldots, T$:

$$\begin{aligned} P_T^t &\equiv \left[P_{\text{GL}}^t P_{\text{GP}}^t\right]^{1/2} \\ &\approx \left[P_L^t P_P^t\right]^{1/2} \varepsilon\left(p^1, p^t, s^1, s^t\right) \\ &= P_F^t \, \varepsilon\left(p^1, p^t, s^1, s^t\right) \end{aligned} \tag{11.41}$$

where the approximation error function $\varepsilon\left(p^1, p^t, s^1, s^t\right)$ is defined as follows for $t = 1, \ldots, T$:

$$\varepsilon\left(p^1, p^t, s^1, s^t\right) \equiv \frac{\left\{1 - \frac{1}{2} \sum_{n=1}^{N} s_{1n} \left[\frac{p_{tn}}{p_{1n} P_L^t} - 1\right]^2\right\}^{\frac{1}{2}}}{\left\{1 - \frac{1}{2} \sum_{n=1}^{N} s_{tn} \left[\frac{p_{1n} P_P^t}{p_{tn}} - 1\right]^2\right\}^{\frac{1}{2}}}. \tag{11.42}$$

Thus P_T^t is approximately equal to P_F^t for $t = 1, \ldots, T$. But how good are these approximations? We know from Diewert (1978) that $P_T^t = P_T\left(p^1, p^t, s^1, s^t\right)$ approximates $P_F^t = P_F\left(p^1, p^t, s^1, s^t\right)$ to the second order around any point where $p^t = p^1$ and $s^t = s^1$.[52] Since the approximations in (11.33) and (11.38) are also second-order approximations, it is likely that the approximation given by (11.41) is fairly good.[53]

In general, if the products are highly substitutable and if prices and shares trend in opposite directions, then we expect that the base period share weighted variance $\sum_{n=1}^{N} s_{1n}\left[\left(p_{tn}/p_{1n} P_L^t\right) - 1\right]^2$ and the current

[52] This result can be generalized to the case where $p^t = \lambda p^1$ and $s^t = s^1$.

[53] However, the Diewert (1978) second-order approximation is different from the present second-order approximations that are derived from Proposition 2. Thus the closeness of $\varepsilon\left(p^1, p^t, s^1, s^t\right)$ to 1 depends on the closeness of the Diewert second-order approximation of P_T^t to P_F^t and the closeness of the second-order approximations that were used in (11.33) and (11.38), which use different Taylor series approximations. Vartia and Suoperä (2018) used alternative Taylor series approximations to obtain relationships between various indexes.

period share weighted variance $\sum_{n=1}^{N} s_{tn}\left[\left(p_{1n}P_P^t/p_{tn}\right)-1\right]^2$ will increase as t increases. It appears that the second variance expression increases more than the first one because the change in expenditure shares from s_{1n} to s_{tn} tends to magnify the squared differences $\left[\left(p_{1n}P_P^t/p_{tn}\right)-1\right]^2$. Thus as say p_{tn} increases and the difference $\left(p_{1n}P_P^t/p_{tn}\right)-1$ decreases, the share s_{tn} will become smaller, and this decreasing share weight s_{tn} will lead to a further shrinkage of the term $s_{tn}\left[\left(p_{1n}P_P^t/p_{tn}\right)-1\right]^2$. On the other hand, if p_{tn} decreases substantially, the difference $\left(p_{1n}P_P^t/p_{tn}\right)-1$ will substantially increase and the share s_{tn} will become larger, and this increasing share weight s_{tn} will further magnify the term $s_{tn}\left[\left(p_{1n}P_P^t/p_{tn}\right)-1\right]^2$. For large changes in prices, the magnification effects will tend to be more important than the shrinkage effects of changing expenditure shares. This overall *share magnification effect* does not occur for the base period share weighted variance $\sum_{n=1}^{N} s_{1n}\left[\left(p_{tn}/p_{1n}P_L^t\right)-1\right]^2$. Thus if the products are highly substitutable and there are large divergent trends in prices, P_T will tend to increase relative to P_F as time increases under these conditions. The more substitutable the products are, the greater will be this tendency.

Our tentative conclusion at this point is that the approximations defined by (11.33), (11.38) and (11.41) are good enough to provide rough estimates of the differences in the six price indexes involved in these approximate equalities. In an empirical example using scanner data, Diewert (2018) found that the variance terms on the right hand sides of (11.38) tended to be larger than the corresponding variances on the right hand sides of (11.33) and these differences led to a tendency for the fixed base Fisher price indexes P_F^t to be slightly smaller than the corresponding fixed base Törnqvist Theil price indexes P_T^t.[54]

We conclude this section by developing an exact relationship between the geometric Laspeyres and Paasche price indexes. Using definitions (11.32) and (11.35) for the logarithms of these indexes, we have the following exact decomposition for the logarithmic difference between

[54] Vartia and Suoperä (2018) also found a tendency for the Fisher price index to lie slightly below their Törnqvist counterparts in their empirical work.

these indexes for $t = 1, \ldots, T$[55]:

$$\ln P_{\mathrm{GP}}^t - \ln P_{\mathrm{GL}}^t = \sum_{n=1}^{N} s_{tn} \ln\left(\frac{p_{tn}}{p_{1n}}\right) - \sum_{n=1}^{N} s_{1n} \ln\left(\frac{p_{tn}}{p_{1n}}\right)$$
$$= \sum_{n=1}^{N} [s_{tn} - s_{1n}][\ln p_{tn} - \ln p_{1n}]. \quad (11.43)$$

Define the vectors $\ln p^t \equiv [\ln p_{t1}, \ln p_{t2}, \ldots, \ln p_{tN}]$ for $t = 1, \ldots, T$. It can be seen that the right hand side of equation t in (11.43) is equal to $\left[s^t - s^1\right]\left[\ln p^t - \ln p^1\right]$, the inner product of the vectors $x \equiv s^t - s^1$ and $y \equiv \ln p^t - \ln p^1$. Let x^* and y^* denote the arithmetic means of the components of the vectors x and y. Note that $x^* \equiv (1/N)1_N x = (1/N)1_N \cdot \left[s^t - s^1\right] = (1/N)[1 - 1] = 0$. The covariance between x and y is defined as $\mathrm{cov}(x, y) \equiv (1/N)\left[x - x^* 1_N\right] \cdot \left[y - y^* 1_N\right] = (1/N)x \cdot y - x^* y^* = (1/N)x \cdot y$[56] since x^* is equal to 0. Thus the right hand side of (11.43) is equal to $N\mathrm{cov}(x, y) = N\mathrm{cov}\left(s^t - s^1, \ln p^t - \ln p^1\right)$; i.e., the right hand side of (11.43) is equal to N times the covariance of the long-term share difference vector, $s^t - s^1$, with the long-term log price difference vector, $\ln p^t - \ln p^1$. Hence, if this covariance is positive, then $\ln P_{\mathrm{GP}}^t - \ln P_{\mathrm{GL}}^t > 0$ and $P_{\mathrm{GP}}^t > P_{\mathrm{GL}}^t$. If this covariance is negative, then $P_{\mathrm{GP}}^t < P_{\mathrm{GL}}^t$. We argue below that for the case where the N products are close substitutes, it is likely that the covariances on the right hand side of Eq. (11.43) are negative for $t > 1$.

Suppose that the observed price and quantity data are approximately consistent with purchasers having identical Constant Elasticity of Substitution preferences. CES preferences are dual to the CES unit cost function $m_{r,\alpha}(p)$, which is defined by (11.2) above, where α satisfies (11.1) and $r \leq 1$. It can be shown[57] that the sales share for product n in a period where purchasers face the strictly positive price vector

[55] Vartia and Suoperä (2018, 26) derived this result and noticed that the right hand side of (11.43) could be interpreted as a covariance. They also developed several alternative exact decompositions for the difference $\ln P_{\mathrm{GP}}^t - \ln P_{\mathrm{GL}}^t$. Their paper also develops a new theory of "excellent" index numbers.

[56] This equation is the *covariance identity* that was first used by Bortkiewicz (1923) to show that normally the Paasche price index is less than the corresponding Laspeyres index.

[57] See Eq. (11.110) in Diewert (2021a) or Diewert and Feenstra (2017).

$p \equiv [p_1, \ldots, p_N]$ is the following share:

$$s_n(p) \equiv \frac{\alpha_n p_n^r}{\sum_{i=1}^{N} \alpha_i p_i^r}; \quad n = 1, \ldots, N. \tag{11.44}$$

Upon differentiating $s_n(p)$ with respect to p_n, we find that the following relations hold:

$$\frac{\partial \ln s_n(p)}{\partial \ln p_n} = r[1 - s_n(p)], \quad n = 1, \ldots, N. \tag{11.45}$$

Thus $\partial \ln s_n(p) / \partial \ln p_n < 0$ if $r < 0$ (or equivalently, if the elasticity of substitution $\sigma \equiv 1 - r$ is greater than 1) and $\partial \ln s_n(p) / \partial \ln p_n > 0$ if r satisfies $0 < r < 1$ (or equivalently, if the elasticity of substitution satisfies $0 < \sigma < 1$).[58] If we are aggregating prices at the first stage of aggregation where the products are close substitutes and purchasers have common CES preferences, then it is likely that the elasticity of substitution is greater than 1 and hence as the price of product n decreases, it is likely that the share of that product will increase. Hence, we expect the terms $[s_{tn} - s_{1n}][\ln p_{tn} - \ln p_{1n}]$ to be predominantly negative; i.e., if p_{1n} is unusually low, then $\ln p_{tn} - \ln p_{1n}$ is likely to be positive and $s_{tn} - s_{1n}$ is likely to be negative. On the other hand, if p_{tn} is unusually low, then $\ln p_{tn} - \ln p_{1n}$ is likely to be negative and $s_{tn} - s_{1n}$ is likely to be positive. Thus for closely related products, we expect the covariances on the right hand sides of (11.43) to be negative and for P_{GP}^t to be less than P_{GL}^t. We can combine this inequality with our previously established inequalities to conclude that for closely related products, it is likely that $P_P^t < P_{\text{GP}}^t < P_T^t < P_{\text{GL}}^t < P_L^t$. On the other hand, if we are aggregating at higher levels of aggregation, then it is likely that the elasticity of substitution is in the range $0 < \sigma < 1$,[59] and in this case, the covariances on the right hand sides of (11.43) will tend to be positive and hence in this

[58] Thus define product n to be a *strong substitute* with all other products if $\partial \ln s_n(p) / \partial \ln p_n < 0$ and to be a *weak substitute* if $\partial \ln s_n(p) / \partial \ln p_n > 0$.

[59] See Shapiro and Wilcox (1997) who found that $\sigma > 0.7$ fit the US data well at higher levels of aggregation. See also Armknecht and Silver (2014, 9) who noted that estimates for σ tend to be greater than 1 at the lowest level of aggregation and less than 1 at higher levels of aggregation.

case, it is likely that $P^t_{\text{GP}} > P^t_{\text{GL}}$. We also have the inequalities $P^t_P < P^t_{\text{GP}}$ and $P^t_{\text{GL}} < P^t_L$ in this case.[60]

We turn now to some relationships between weighted and unweighted (i.e., equally weighted) geometric price indexes.

Relationships Between the Jevons, Geometric Laspeyres, Geometric Paasche and Törnqvist Price Indexes

In this section, we will investigate how close the unweighted Jevons index P^t_J is to the geometric Laspeyres P^t_{GL}, geometric Paasche P^t_{GP} and Törnqvist P^t_T price indexes.

We first investigate the difference between the logarithms of P^t_{GL} and P^t_J. Using the definitions for these indexes, we have the following log differences for $t = 1, \ldots, T$:

$$\begin{aligned} \ln P^t_{\text{GL}} - \ln P^t_J &= \sum_{n=1}^{N} \left[s_{1n} - \left(\frac{1}{N}\right) \right] [\ln p_{tn} - \ln p_{1n}] \\ &= N \operatorname{cov}\left(s^1 - \left(\frac{1}{N}\right) 1_N,\ \ln p^t - \ln p^1 \right) \\ &\equiv \varepsilon_t . \end{aligned} \tag{11.46}$$

In the elementary index context where the N products are close substitutes and product shares in period 1 are close to being equal, it is likely that ε_t is positive; i.e., $\ln p_{1n}$ is unusually low, then s_{1n} is likely to be unusually high and thus it is likely that $s_{1n} - (1/N) > 0$ and $\ln p_{tn} - \ln p_{1n}$ minus the mean of the log ratios $\ln(p_{tn}/p_{1n})$ is likely to be greater than 0 and hence ε_t is likely to be greater than 0, implying that $P^t_{\text{GL}} > P^t_J$. However, if N is small and the shares have a high variance and if product n goes on sale in period 1, then we cannot assert that s_{1n} is likely to be greater than $1/N$ and hence we cannot be confident that ε_t is likely to be

[60] See Vartia (1978, 276–290) for a similar discussion about the relationships between P^t_L, P^t_P, P^t_F, P^t_{GL}, P^t_{GP} and P^t_T. Vartia extended the discussion to include period 1 and period t share weighted harmonic averages of the price ratios, p_{tn}/p_{1n}. See also Armknecht and Silver (2014, 10) for a discussion on how weighted averages of the above indexes could approximate a superlative index at higher levels of aggregation.

greater than 0 and hence we cannot predict with certainty that P_{GL}^t will be greater than P_J^t.

There are three simple sets of conditions that will imply that $P_{GL}^t = P_J^t$: (i) the covariance on the right hand side of (11.46) equals 0; i.e., $\text{cov}(s^1 - (1/N)1_N, \ln p^t - \ln p^1) = 0$; (ii) period t price proportionality; i.e., $p^t = \lambda_t p^1$ for some $\lambda_t > 0$; (iii) equal sales shares in period 1; i.e., $s^1 = (1/N)1_N$.

Now look at the difference between the logarithms of P_{GL}^t and P_J^t. Using the definitions for these indexes, for $t = 1, \ldots, T$, we have:

$$\begin{aligned} \ln P_{GP}^t - \ln P_J^t &= \sum_{n=1}^{N} \left[s_{tn} - \frac{1}{N} \right] [\ln p_{tn} - \ln p_{1n}] \\ &= N\text{cov}\left(s^t - \frac{1}{N} 1_N, \ln p^t - \ln p^1 \right) \\ &\equiv \eta_t. \end{aligned} \tag{11.47}$$

In the elementary index context where the N products are close substitutes and the shares s^t are close to being equal, then it is likely that η_t is negative; i.e., if $\ln p_{tn}$ is unusually low, then s_{tn} is likely to be unusually high and thus it is likely that $s_{tn} - (1/N) > 0$ and $\ln p_{tn} - \ln p_{1n}$ minus the mean of the log ratios $\ln(p_{tn}/p_{1n})$ is likely to be less than 0 and hence η_t is likely to be less than 0 implying that $P_{GP}^t < P_J^t$. However, if N is small and the period t shares s^t are not close to being equal, then again, we cannot confidently predict the sign of the covariance in (11.47).

Again, there are three simple sets of conditions that will imply that $P_{GP}^t = P_J^t$: (i) the covariance on the right hand side of (11.47) equals 0; i.e., $\text{cov}(s^t - (1/N)1_N, \ln p^t - \ln p^1) = 0$; (ii) period t price proportionality; i.e., $p^t = \lambda_t p^1$ for some $\lambda_t > 0$; (iii) equal sales shares in period t; i.e., $s^t = (1/N)1_N$.

Using the definitions for P_T^t and P_J^t, the log difference between these indexes is equal to the following expression for $t = 1, \ldots, T$:

$$\begin{aligned} \ln P_T^t - \ln P_J^t &= \sum_{n=1}^{N} \left[\frac{1}{2} s_{tn} + \frac{1}{2} s_{1n} - \frac{1}{N} \right] [\ln p_{tn} - \ln p_{1n}] \\ &= N\text{cov}\left(\frac{1}{2} s^t + \frac{1}{2} s^1 - \frac{1}{N} 1_N, \ln p^t - \ln p^1 \right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{N}{2}\operatorname{cov}\left(s^t - \frac{1}{N}1_N, \ln p^t - \ln p^1\right) \\
&\quad + \frac{N}{2}\operatorname{cov}\left(s^1 - \frac{1}{N}1_N, \ln p^t - \ln p^1\right) \\
&= \frac{1}{2}\varepsilon_t + \frac{1}{2}\eta_t. \qquad (11.48)
\end{aligned}
$$

As usual, there are three simple sets of conditions that will imply that $P_T^t = P_J^t$: (i) the covariance on the right hand side of (11.48) equals 0; i.e., $\operatorname{cov}((1/2)s^t + (1/2)s^1 - (1/N)1_N, \ln p^t - \ln p^1) = 0 = (1/2)\varepsilon_t + (1/2)\eta_t$ or equivalently, $\operatorname{cov}(s^t - (1/N)1_N, \ln p^t - \ln p^1) = -\operatorname{cov}(s^1 - (1/N)1_N, \ln p^t - \ln p^1)$; (ii) period t price proportionality; i.e., $p^t = \lambda_t p^1$ for some $\lambda_t > 0$; (iii) the arithmetic average of the period 1 and t sales shares are all equal to $1/N$; i.e., $(1/2)s^t + (1/2)s^1 = (1/N)1_N$.

If the trend deflated prices p_{tn}/λ_t are distributed *independently* across time and *independently* of the sales shares s_{tn}, then it can be seen that the expected values of the ε_t and η_t will be 0 and hence $P_T^t \approx P_J^t$ for $t = 1, \ldots, T$. Thus it can be the case that the ordinary Jevons price index is able to provide an adequate approximation to the superlative Törnqvist price index in the elementary price index context. However, if the shares are trending and if prices are trending in divergent directions, then P_J^t will not be able to approximate P_T^t.

In the general case, we expect P_T^t to be less than P_J^t. The mean of the average shares for product n in periods 1 and t, $(1/2)s_{tn} + (1/2)s_{1n}$, is $1/N$. Define the means of the log prices in period t as $\ln p_t^* \equiv (1/N)\sum_{n=1}^{N} \ln p_{tn}$ for $t = 1, \ldots, T$. Note that p_t^* is the geometric mean of the period t prices. Thus using the first line of (11.48) and the covariance identity, we have:

$$
\begin{aligned}
\ln P_T^t - \ln P_J^t &= \sum_{n=1}^{N}\left[\frac{1}{2}s_{tn} + \frac{1}{2}s_{1n} - \frac{1}{N}\right][\ln p_{tn} - \ln p_{1n}] \\
&= \sum_{n=1}^{N}\left[\frac{1}{2}s_{tn} + \frac{1}{2}s_{1n} - \frac{1}{N}\right][\ln p_{tn} - \ln p_{1n} - \ln p_t^* + \ln p_1^*] \\
&= \sum_{n=1}^{N}\left[\frac{1}{2}s_{tn} + \frac{1}{2}s_{1n} - \frac{1}{N}\right]\left[\ln\left(\frac{p_{tn}}{p_t^*}\right) - \ln\left(\frac{p_{1n}}{p_1^*}\right)\right] \qquad (11.49)
\end{aligned}
$$

The second line in (11.49) follows from the first line because $\sum_{n=1}^{N}[(1/2)s_{tn} + (1/2)s_{1n} - (1/N)] = 0$ so if these N terms are multiplied by a constant, the resulting sum of terms will still equal 0. Define the *deflated price* for product n in period t as p_{tn}/p_t^* for $t = 1, \ldots, T$. Assume that the products are highly substitutable. Suppose that the deflated price of product n goes down between periods 1 and t so that $\ln(p_{tn}/p_t^*) - \ln(p_{1n}/p_1^*)$ is negative. Under these conditions, there will be a tendency for the average expenditure share for product n, $(1/2)s_{tn} + (1/2)s_{1n}$, to be greater than the average of these shares, which is $1/N$. Thus the term $[(1/2)s_{tn} + (1/2)s_{1n} - (1/N)][\ln(p_{tn}/p_t^*) - \ln(p_{1n}/p_1^*)]$ is *likely* to be negative. Now suppose that the deflated price of product n goes up between periods 1 and t so that $\ln(p_{tn}/p_t^*) - \ln(p_{1n}/p_1^*)$ is positive. Under these conditions, there will be a tendency for the average expenditure share for product n, $(1/2)s_{tn} + (1/2)s_{1n}$, to be less than the average of these shares. Again, the term $[(1/2)s_{tn} + (1/2)s_{1n} - (1/N)][\ln(p_{tn}/p_t^*) - \ln(p_{1n}/p_1^*)]$ is *likely* to be negative. Thus if the products under consideration are highly substitutable, we expect P_T^t to be less than P_J^t.[61] If the products are not highly substitutable, we expect P_T^t to be greater than P_J^t.

The results in this section can be summarized as follows: the unweighted Jevons index, P_J^t, can provide a reasonable approximation to a fixed base superlative index like P_T^t *provided* that the expenditure shares do not systematically trend with time and prices do not systematically grow at diverging rates. If these assumptions are not satisfied, then it is likely that the Jevons index will have some bias relative to a superlative index; P_J^t is likely to exceed P_T^t as t becomes large if the products are close substitutes and P_J^t is likely to be less than P_T^t if the products are not close substitutes.

[61] This is perhaps an important result in the context where a statistical agency is collecting web scraped prices for very similar products and using an equally weighted geometric mean of these scraped prices as an estimated elementary price level. The resulting Jevons price index may have an upward bias relative to its superlative counterpart.

Relationships Between Superlative Fixed Base Indexes and Geometric Indexes That Use Average Annual Shares as Weights

We consider the properties of weighted Jevons indexes where the weight vector is an *annual average* of the observed monthly shares in a previous year. Recall that the weighted Jevons (or Cobb Douglas) price index $P_{J\alpha}^t$ was defined by (11.15) in section "Comparing CES Price Levels and Price Indexes" as $P_{J\alpha}^t \equiv \prod_{n=1}^N \left(p_{tn}/p_{1n}\right)^{\alpha_n}$ where the product weighting vector α satisfied the restrictions $\alpha >> 0_N$ and $\alpha \cdot 1_N = 1$. The following counterparts to the covariance identities (11.46)–(11.48) hold for $t = 1, \ldots,$ T where the Geometric Young index or weighted Jevons index $P_{J\alpha}^t$ has replaced P_J^t.[62]

$$\ln P_{\mathrm{GL}}^t - \ln P_{J\alpha}^t = \sum_{n=1}^N (s_{1n} - \alpha_n)(\ln p_{tn} - \ln p_{1n}) = N\mathrm{cov}\left(s^1 - \alpha, \ln p^t - \ln p^1\right); \quad (11.50)$$

$$\ln P_{\mathrm{GP}}^t - \ln P_{J\alpha}^t = \sum_{n=1}^N (s_{tn} - \alpha_n)(\ln p_{tn} - \ln p_{1n}) = N\mathrm{cov}\left(s^t - \alpha, \ \ln p^t - \ln p^1\right); \quad (11.51)$$

$$\begin{aligned} \ln P_T^t - \ln P_{J\alpha}^t &= \sum_{n=1}^N \left[\frac{1}{2}s_{tn} + \frac{1}{2}s_{1n} - \alpha_n\right][\ln p_{tn} - \ln p_{1n}] \\ &= N\mathrm{cov}\left[\frac{1}{2}s^t + \frac{1}{2}s^1 - \alpha, \ \ln p^t - \ln p^1\right] \\ &= \frac{1}{2}\left[\ln P_{\mathrm{GL}}^t - \ln P_{J\alpha}^t\right] + \frac{1}{2}\left[\ln P_{\mathrm{GP}}^t - \ln P_{J\alpha}^t\right]. \end{aligned} \quad (11.52)$$

[62] The relationship (11.52) was obtained by Armknecht and Silver (2014, 9); i.e., take logarithms on both sides of their Eq. (11.12) and we obtain the first equation in Eq. (11.52).

Define α as the *arithmetic average of the first T^* observed share vectors s^t*:

$$\alpha = \sum_{t=1}^{T^*} \frac{1}{T^*} s^t. \tag{11.53}$$

In the context where the data consists of monthly periods, T^* will typically be equal to 12; i.e., the elementary index under consideration is the weighted Jevons index $P_{J\alpha}^t$ where the weight vector α is the average of the observed expenditure shares for the first 12 months in the sample.

The decompositions (11.50)–(11.52) will hold for the α defined by (11.53). If the N products are highly substitutable, it is likely that $\text{cov}(s^1 - \alpha, \ln p^t - \ln p^1) > 0$ and $\text{cov}(s^t - \alpha, \ln p^t - \ln p^1) < 0$ and hence, it is likely that $P_{\text{GL}}^t > P_{J\alpha}^t$ and $P_{\text{GP}}^t < P_{J\alpha}^t$. If the products are not close substitutes, then it is likely that $P_{\text{GL}}^t < P_{J\alpha}^t$ and $P_{\text{GP}}^t > P_{J\alpha}^t$. If there are no divergent trends in prices, then it is possible that the *average share price index* $P_{J\alpha}^t$ could provide an adequate approximation to the superlative Törnqvist index P_T^t.

Note that t takes on the values $t = 1, \ldots, T$ in Eqs. (11.50)–(11.52). However, annual share indexes that are implemented by statistical agencies are not constructed in exactly this manner. The practical month-to-month indexes that are constructed by statistical agencies using annual shares of the type defined by (11.53) do not choose the reference month for prices to be month 1; rather, they chose the reference month for prices to be $T^* + 1$, the month that follows the first year.[63] Thus the reference *year* for share weights precedes the reference *month* for prices. In this case, the logarithm of the month $t \geq T^* + 1$ annual share weighted Jevons index, $\ln P_{J\alpha}^t$, is defined as follows where α is the vector of annual average share weights defined by (11.53):

$$\ln P_{J\alpha}^t \equiv \sum_{n=1}^{N} \alpha_n (\ln p_{tn} - \ln p_{T^*+1,n}); \quad t = T^* + 1,\ T^* + 2, \ldots, T. \tag{11.54}$$

The following counterparts to the identities (11.50)–(11.52) hold for $t = T^* + 1, T^* + 2, \ldots, T$ where α is defined by (11.53) and $P_{J\alpha}^t$ is defined

[63] In actual practice, the reference month for prices can be many months after T^*.

by (11.54):

$$\ln P_{\mathrm{GL}}^t - \ln P_{J\alpha}^t = \sum_{n=1}^{N} \left(s_{T^*+1,n} - \alpha_n\right)\left(\ln p_{tn} - \ln p_{T^*+1,n}\right)$$
$$= N\mathrm{cov}\left(s^{T^*+1} - \alpha,\ \ \ln p^t - \ln p^{T^*+1}\right); \qquad (11.55)$$

$$\ln P_{GP}^t - \ln P_{J\alpha}^t = \sum_{n=1}^{N} (s_{tn} - \alpha_n)\left(\ln p_{tn} - \ln p_{T^*+1,n}\right)$$
$$= N\mathrm{cov}\left(s^t - \alpha, \ln p^t - \ln p^{T^*+1}\right); \qquad (11.56)$$

$$\ln P_T^t - \ln P_{J\alpha}^t = \sum_{n=1}^{N} \left(\frac{1}{2}s_{tn} + \frac{1}{2}s_{T^*+1,n} - \alpha_n\right)\left(\ln p_{tn} - \ln p_{T^*+1,n}\right)$$
$$= N\mathrm{cov}\left(\frac{1}{2}s^t + \frac{1}{2}s^{T^*+1} - \alpha,\ \ \ln p^t - \ln p^{T^*+1}\right)$$
$$= \frac{1}{2}\left(\ln P_{\mathrm{GL}}^t - \ln P_{J\alpha}^t\right) + \frac{1}{2}\left(\ln P_{\mathrm{GP}}^t - \ln P_{J\alpha}^t\right). \qquad (11.57)$$

If the N products are highly substitutable, it is likely that $\mathrm{cov}\left(s^{T^*+1} - \alpha,\ \ln p^t - \ln p^{T^*+1}\right) > 0$ so that $P_{\mathrm{GL}}^t > P_{J\alpha}^t$. It is also likely that $\mathrm{cov}\left(s^t - \alpha, \ln p^t - \ln p^{T^*+1}\right) < 0$ and hence it is likely that $P_{\mathrm{GP}}^t < P_{J\alpha}^t$ in the highly substitutable case. If the products are not close substitutes, then it is likely that $P_{\mathrm{GL}}^t < P_{J\alpha}^t$ and $P_{\mathrm{GP}}^t > P_{J\alpha}^t$. If there are no divergent trends in prices, then it is possible that the *average share price index* $P_{J\alpha}^t$ could provide an adequate approximation to the superlative Törnqvist index P_T^t. However, if there are divergent trends in prices and shares and the products are highly substitutable with each other, then we expect the covariance in (11.56) to be more negative than the covariance in (11.55) is positive so that P_T^t will tend to be less than the annual shares geometric index $P_{J\alpha}^t$. Thus $P_{J\alpha}^t$. will tend to have a bit of substitution bias if the products are highly substitutable, which is an intuitively plausible result.

As usual, there are three simple sets of conditions that will imply that $P_T^t = P_{J\alpha}^t$: (i) the covariance on the right hand side of (11.57) equals 0; i.e., $\mathrm{cov}\left[(1/2)s^t + (1/2)s^{T^*+1} - \alpha,\ \ \ln p^t - \ln p^{T^*+1}\right] =$

0 or equivalently, $\text{cov}\left(s^{T^*+1} - \alpha, \ln p^t - \ln p^{T^*+1}\right) = -\text{cov}\left(s^t - \alpha, \ln p^t - \ln p^{T^*+1}\right)$; (ii) period t price proportionality (to the prices of the price reference period); i.e., $p^t = \lambda_t p^{T^*+1}$ for some $\lambda_t > 0$, (iii) the arithmetic average of the period $T^* + 1$ and t sales shares are all equal to α defined by (11.53); i.e., $(1/2)s^t + (1/2)s^{T^*+1} = \alpha$. This last condition will hold if the shares s^t are constant over all time periods and α is defined by (11.53).

Suppose that there are *linear trends in shares* and *divergent linear trends in log prices*; i.e., suppose that the following assumptions hold for $t = 2, 3, , \ldots, T$:

$$s^t = s^1 + \beta(t - 1); \tag{11.58}$$

$$\ln p^t = \ln p^1 + \gamma(t - 1) \tag{11.59}$$

where $\beta \equiv [\beta_1, \ldots, \beta_N]$ and $\gamma \equiv \left[\gamma_1, \ldots, \gamma_N\right]$ are constant vectors and β satisfies the additional restriction[64]:

$$\beta \cdot 1_N = 0. \tag{11.60}$$

In the case where the products are highly substitutable, if the price of product n, p_{tn}, is trending upwards so that γ_n is positive, then we could expect that the corresponding share s_{tn} is trending downward so that β_n is negative. Similarly, if γ_n is negative, then we expect that the corresponding β_n is positive. Thus we expect that $\sum_{n=1}^{N} \beta_n \gamma_n = \beta \cdot \gamma < 0$.

Substituting (11.58) into definition (11.53) gives us the following expression for the annual share weight vector under the linear trends assumption:

$$\begin{aligned} \alpha \equiv \sum_{t=1}^{T^*} \frac{1}{T^*} s^t &= \sum_{t=1}^{T^*} \frac{1}{T^*}\left[s^1 + \beta(t-1)\right]. \\ &= s^1 + \frac{1}{2}\beta\left(T^* - 1\right). \end{aligned} \tag{11.61}$$

[64] Since expenditure shares must be nonnegative, if $\beta \neq 0_N$ then some components of β will be negative and thus the linear trends in shares assumption (11.58) cannot hold forever. Assumptions (11.58) and (11.59) will generally be only approximately true and they cannot hold indefinitely.

Using (11.57)–(11.59) and (11.61), we have the following equations for $t = T^* + 1, T^* + 2, \ldots, T$:

$$\ln P_T^t - \ln P_{J\alpha}^t = \left[\frac{1}{2}s^t + \frac{1}{2}s^1 - \alpha\right]\left[\ln p^t - \ln p^1\right] = \frac{1}{2}\beta \cdot \gamma t\left(t - T^* - 1\right). \quad (11.62)$$

.

Thus if the inner product of the vectors β and γ is not equal to 0, $\ln P_T^t$ and $\ln P_{J\alpha}^t$ will diverge at a *quadratic rate* as t increases. Under these trend assumptions, the average share geometric index $P_{J\alpha}^t$ will be subject to some substitution bias (as compared to P_T^t which controls for substitution bias[65]), which will grow over time.[66] As indicated above, it is likely that $\beta \cdot \gamma < 0$ so that it is likely that P_T^t will be below $P_{J\alpha}^t$ under the assumption of strong substitutability and diverging trends in prices and shares.

Note that in real life, new products appear and existing products disappear. The analysis presented in this section and in previous sections can take this fact into account *in theory* if the price statistician has somehow calculated approximate reservation prices for products that are not available in the current period. Note that product churn means that shares are not constant over time; i.e., *product churn will lead to nonsmooth trends in product shares*. However, superlative indexes like ${P_F}^t$ and ${P_T}^t$ can deal with new and disappearing products in a way that is consistent with consumer theory, provided that suitable reservation prices have been either estimated or approximated by suitable rules of thumb.

To Chain or Not to Chain

In the above discussions, attention has been focused on direct indexes that compare the prices of period t with the prices of period 1. But it is also

[65] We regard an index as having some substitution bias if it diverges from a superlative index which controls for substitution bias. See Diewert (1976) for the formal definition of a superlative index.

[66] If all prices grow at the same geometric rate, then it can be verified that $P_{J\alpha}^t = P_{\text{GL}}^t = P_{\text{GP}}^t = P_T^t$. If in addition, assumptions (11.58)–(11.60) hold, then $\gamma = \lambda 1_N$ for some scalar $\lambda > 0$ and using assumption (11.60), we have $\beta \cdot \gamma = 0$ and thus $P_T^t = P_{J\alpha}^t$ under our assumptions.

possible to move from period 1 prices to period t prices by moving from one period to the next and cumulating the jumps. If the second method is used, the resulting period t price index is called a *chained index*. In this section, we will examine the possible differences between direct and chained Törnqvist price indexes.

It is convenient to introduce some new notation. Denote the Törnqvist price index that compares the prices of period j to the prices of period i (the base period for the comparison) by $P_T(i, j)$. The logarithm of $P_T(i, j)$ is defined as follows for $i, j = 1, \ldots, T$:

$$\begin{aligned}\ln P_T(i, j) &\equiv \frac{1}{2}\sum_{n=1}^{N}\left(s_{in} + s_{jn}\right)\left(\ln p_{jn} - \ln p_{in}\right)\\ &= \frac{1}{2}\left(s^i + s^j\right)\left(\ln p^j - \ln p^i\right).\end{aligned} \tag{11.63}$$

The chained Törnqvist price index going from period 1 to T will coincide with the corresponding direct index if the indexes $P_T(i, j)$ satisfy the following *multiperiod identity test*, which is due to Walsh (1901, 389; 1921b, 540):

$$P_T(1, 2)P_T(2, 3)\ldots P_T(T-1, T)P_T(T, 1) = 1. \tag{11.64}$$

The above test can be used to measure the amount that the chained indexes between periods 1 and T differ from the corresponding direct index that compares the prices of period 1 and T; i.e., if the product of indexes on the left hand side of (11.64) is different from unity, then we say that the index number formula is subject to *chain drift* and the difference between the left and right hand sides of (11.64) serves to measure the magnitude of the chain drift problem.[67] In order to determine whether the Törnqvist price index formula satisfies the multiperiod identity test (11.64), take the logarithm of the left hand side of (11.64) and check whether it is equal to the logarithm of 1 which is 0. Thus substituting definitions (11.63) into the logarithm of the left hand side

[67] Walsh (1901, 401) was the first to propose this methodology to measure chain drift. It was independently proposed later by Persons (1921, 110) and Szulc (1983, 540). Fisher's (1922, 284) circular gap test could also be interpreted as a test for chain drift.

of (11.64) leads to the following expressions[68]:

$$
\begin{aligned}
&\ln P_T(1, 2) + \ln P_T(2, 3) + \ldots + \ln P_T(T-1, T) + \ln P_T(T, 1) \\
&\quad = \frac{1}{2}\sum_{n=1}^{N}(s_{1n} + s_{2n})(\ln p_{2n} - \ln p_{1n}) \\
&\qquad + \frac{1}{2}\sum_{n=1}^{N}(s_{2n} + s_{3n})(\ln p_{3n} - \ln p_{2n}) + \ldots \\
&\qquad + \frac{1}{2}\sum_{n=1}^{N}\left(s_{T-1,n} + s_{Tn}\right)\left(\ln p_{Tn} - \ln p_{T-1,n}\right) \\
&\qquad + \frac{1}{2}\sum_{n=1}^{N}(s_{Tn} + s_{1n})(\ln p_{1n} - \ln p_{Tn}) \\
&\quad = \frac{1}{2}\sum_{n=1}^{N}(s_{1n} - s_{3n})\ln p_{2n} + \frac{1}{2}\sum_{n=1}^{N}(s_{2n} - s_{4n})\ln p_{3n} \\
&\qquad + \ldots + \frac{1}{2}\sum_{n=1}^{N}\left(s_{T-2,n} - s_{Tn}\right)\ln p_{T-1,n} \\
&\qquad + \frac{1}{2}\sum_{n=1}^{N}(s_{Tn} - s_{2n})\ln p_{1n} + \frac{1}{2}\sum_{n=1}^{N}\left(s_{T-1,n} - s_{1n}\right)\ln p_{Tn}. \qquad (11.65)
\end{aligned}
$$

In general, it can be seen that the Törnqvist price index formula will be subject to some chain drift; i.e., the sums of terms on the right hand side of (11.65) will not equal 0 in general. However, there are four sets of conditions where these terms will sum to 0.

The first set of conditions makes use of the first equality on the right hand side of (11.65). If the prices vary in strict proportion over time, so that $p^t - \lambda_t p^1$ for $t = 2, 3, \ldots, T$, then it is straightforward to show that (11.64) is satisfied.

68 Persons (1928, 101) developed a similar decomposition using the bilateral Fisher formula instead of the Törnqvist formula. See also de Haan and Krsinich (2014) for an alternative decomposition.

The second set of conditions makes use of the second equality in Eq. (11.65). If the shares s^t are constant over time,[69] then it is obvious that (11.64) is satisfied.

The third set of conditions also makes use of the second equality in (11.65). The sum of terms $\sum_{n=1}^{N}(s_{1n} - s_{3n}) \ln p_{2n}$ is equal to $(s^1 - s^3) \ln p^2$ which in turn is equal to $(s^1 - s^3) \cdot (\ln p^2 - \ln p^{2*}) = N\mathrm{cov}(s^1 - s^3, \ln p^2)$ where $\ln p^{2*} \equiv (1/N)\sum_{n=1}^{N} \ln p_{2n}$, the mean of the components of $\ln p^2$. Thus the N sets of summations on the right hand side of the second equation in (11.65) can be interpreted as constants times the covariances of a difference in shares (separated by one or more time periods) with the logarithm of a price vector for a time period that is not equal to either of the time periods involved in the difference in shares. Thus, if the covariance equalities $\mathrm{cov}(s^1 - s^3, \ln p^2) = \mathrm{cov}(s^2 - s^4, \ln p^3) = \ldots = \mathrm{cov}(s^{T-2} - s^T, \ln p^{T-1}) = \mathrm{cov}(s^T - s^2, \ln p^1) = \mathrm{cov}(s^{T-1} - s^1, \ln p^T) = 0$ hold, then (11.64) will be satisfied. These zero covariance conditions will be satisfied if the log prices of one period are uncorrelated with the shares of all other periods. If the time period is long enough and there are no trends in log prices and shares, so that prices are merely bouncing around in a random fashion,[70] then these zero covariance conditions are likely to be satisfied to a high degree of approximation and thus under these conditions, the Törnqvist Theil price index is likely to be largely free of chain drift. However, in the elementary index context where retailers have periodic highly discounted prices, the zero correlation conditions are unlikely to hold. Suppose that product n goes on sale during period 2 so that $\ln p_{2n}$ is well below the average price for period 2. Suppose product n is not on sale during periods 1 and 3. If purchasers have stocked up on product n during period 2, it is likely that s_{3n} will be less than s_{1n} and thus it is likely that $\mathrm{cov}(s^1 - s^3, \ln p^2) < 0$. Now suppose that product n is not on sale during period 2. In this case, it is likely that $\ln p_{2n}$ is greater than the average log price during period 2. If product n was on sale during period 1 but not period 3, then s_{1n} will tend to be greater than s_{3n} and thus $\mathrm{cov}(s^1 - s^3, \ln p^2) > 0$. However,

[69] If purchasers of the products have Cobb-Douglas preferences, then the sales shares will be constant.

[70] Szulc (1983) introduced the term "price bouncing" to describe the behavior of soft drink prices in Canada at the elementary level.

Table 11.3 Prices and quantities for two products and the Fisher fixed base and chained price indexes

t	p_{t1}	p_{t2}	q_{t1}	q_{t2}	P_F^t	P_{FCh}^t
1	2	1	100	1	1.00000	1.00000
2	10	1	40	40	4.27321	4.27321
3	10	1	25	80	3.55553	4.27321
4	5	2	50	20	2.45676	2.96563

if product n was on sale during period 3 but not period 1, then s_{1n} will tend to be less than s_{3n} and thus $\text{cov}(s^1 - s^3, \ln p^2) < 0$. These last two cases should largely offset each other and so we are left with the likelihood that $\text{cov}(s^1 - s^3, \ln p^2) < 0$. Similar arguments apply to the other covariances and so we are left with the expectation that the chained Törnqvist index used in the elementary index context is likely to drift downwards relative to its fixed base counterpart.[71]

Since the Fisher index normally approximates the Törnqvist fairly closely, we expect both the chained Fisher and Törnqvist indexes to exhibit downward chain drift. However, it is not always the case that a superlative index is subject to downward chain drift. Feenstra and Shapiro (2003) found upward chain drift in the Törnqvist formula using a scanner data set. Persons (1928, 100–105) had an extensive discussion of the chain drift problem with the Fisher index and he gave a numerical example on page 102 of his article that showed how upward chain drift could occur. We have adapted his example in Table 11.3.

Product 1 is on sale in period 1 and goes back to a relatively high price in periods 2 and 3 and then goes on sale again but the discount is not as steep as the period 1 discount. Product 2 is at its "regular" price for periods 1–3 and then rises steeply in period 4. Products 1 and 2 are close substitutes so when product 1 is steeply discounted, only 1 unit of product 2 is sold in period 1 while 100 units of product 1 are sold.

[71] Fisher (1922, 284) found little difference in the fixed base and chained Fisher indexes for his particular data set which he used to compare 119 different index number formulae. Fisher noted that the Carli, Laspeyres and share weighted Carli chained indexes showed upward chain drift. However, Persons (1921, 110) showed that the Fisher chained index ended up about 4% lower than its fixed base counterpart for his agricultural data set covering 10 years. This is an early example of the downward chain drift associated with the use of the Fisher index.

When the price of product 1 increases fivefold in period 2, demand for the product falls and purchasers switch to product 2 but the adjustment to the new higher price of product 1 is not complete in period 2: in period 3 (where prices are unchanged from period 2), purchasers continue to substitute away from product 1 and toward product 2. It is this *incomplete adjustment* that causes the chained index to climb above the fixed base index in period 3.[72] Thus it is not always the case that the Fisher index is subject to downward chain drift but we do expect that "normally," this would be the case.

The fourth set of conditions that ensure that there is no chain drift are assumptions (11.58) and (11.59), i.e., the assumption that *shares and log prices have linear trends*. To prove this assertion, substitute these equations into either one of the two right hand side equations in (11.65) and we find that the resulting sum of terms is 0.[73] This result is of some importance at higher levels of aggregation where aggregate prices and quantities are more likely to have smooth trends. If the trends are actually linear, then this result shows that there will be no chain drift if the Törnqvist Theil index number formula is used to aggregate the data.[74] However, when this formula is used at the elementary level when there are frequent fluctuations in prices and quantities, chain drift is likely to occur and thus the use of a fixed base index or a multilateral index is preferred under these conditions.

As was mentioned in the introduction, a main advantage of the chain system is that under conditions where prices and quantities are trending smoothly, chaining will reduce the spread between the Paasche and Laspeyres indexes.[75] These two indexes each provide an asymmetric perspective on the amount of price change that has occurred between the two periods under consideration and it could be expected that a single

[72] Persons (1928, 102) explained that it was *incomplete adjustment* that caused the Fisher chained index to climb above the corresponding fixed base index in his example. Ludwig von Auer (2019) proposed a similar theory.

[73] This result was first established by Alterman et al. (1999, 61–65).

[74] This transitivity property carries over to an *approximate* transitivity property for the Fisher and Walsh index number formulae using the fact that these indexes approximate the Törnqvist Theil index to the second order around an equal price and quantity point; see Diewert (1978) on these approximations.

[75] See Diewert (1978, 895) and Hill (1988) for additional discussion on the benefits and costs of chaining.

point estimate of the aggregate price change should lie between these two estimates. Thus at higher levels of aggregation, the use of either a chained Paasche or Laspeyres index will usually lead to a smaller difference between the two and hence to estimates that are closer to the "truth." However, at lower levels of aggregation, smooth changes in prices and quantities are unlikely to occur.

An alternative to the use of a fixed base index is the use of a *multilateral index*. A problem with the use of a fixed base index is that it depends asymmetrically on the choice of the base period. If the structure of prices and quantities for the base period is unusual and fixed base index numbers are used, then the choice of the base period could lead to "unusual" results. Multilateral indexes treat each period symmetrically and thus avoid this problem. In the following section, we will introduce some possible multilateral indexes that are free of chain drift (within our window of T observations).[76]

Relationships Between the Törnqvist Index and the GEKS and CCDI Multilateral Indexes

It is useful to introduce some additional notation at this point. Denote the Laspeyres, Paasche and Fisher price indexes that compare the prices of period j to the prices of period i (the base period for the comparison) by $P_L(i, j)$, $P_P(i, j)$ and $P_F(i, j)$ respectively. These indexes are defined as follows for $r, \ t = 1, \ldots, T$:

$$P_L(r, t) \equiv \frac{p^t \cdot q^r}{p^r \cdot q^r}; \tag{11.66}$$

$$P_P(r, t) \equiv \frac{p^t \cdot q^t}{p^r \cdot q^t}; \tag{11.67}$$

$$P_F(r, t) \equiv [P_L(r, t) P_P(r, t)]^{1/2}. \tag{11.68}$$

The Fisher indexes have very good axiomatic properties and hence are preferred indexes from the viewpoint of the test or axiomatic approach.[77]

[76] Ivancic et al. (2009, 2011) advocated the use of multilateral indexes adapted to the time series context in order to control chain drift. Balk (1980, 1981) also advocated the use of multilateral indexes in order to address the problem of seasonal commodities.

[77] See Diewert (1992) on the axiomatic properties of the Fisher index.

Obviously, one could choose period 1 as the base period and form the following sequence of price levels relative to period 1: $P_F(1, 1) = 1$, $P_F(1, 2)$, $P_F(1, 3)$, ..., $P_F(1, T)$. But one could also use period 2 as the base period and use the following sequence of price levels: $P_F(2, 1)$, $P_F(2, 2) = 1$, $P_F(2, 3)$, ..., $P_F(2, T)$. Each period could be chosen as the base period and thus we end up with T alternative series of Fisher price levels. Since each of these sequences of price levels is equally plausible, Gini (1931) suggested that it would be appropriate to take the geometric average of these alternative price levels in order to determine the final set of price levels. Thus the *GEKS price levels*[78] for periods $t = 1, 2, \ldots, T$ are defined as follows:

$$p^t_{\text{GEKS}} \equiv \left[\prod_{r=1}^{T} P_F(r, t) \right]^{1/T}. \tag{11.69}$$

Note that all time periods are treated in a symmetric manner in the above definitions. The GEKS price indexes P^t_{GEKS} are obtained by normalizing the above price levels so that the period 1 index is equal to 1. Thus we have the following definitions for P^t_{GEKS} for $t = 1, \ldots, T$:

$$P^t_{\text{GEKS}} \equiv \frac{p^t_{\text{GEKS}}}{p^1_{\text{GEKS}}}. \tag{11.70}$$

It is straightforward to verify that the GEKS price indexes satisfy Walsh's multiperiod identity test which becomes the following test in the present context:

$$\left[\frac{P^2_{\text{GEKS}}}{P^1_{\text{GEKS}}}\right]\left[\frac{P^3_{\text{GEKS}}}{P^2_{\text{GEKS}}}\right]\cdots\left[\frac{P^T_{\text{GEKS}}}{P^{T-1}_{\text{GEKS}}}\right]\left[\frac{P^1_{\text{GEKS}}}{P^T_{\text{GEKS}}}\right] = 1. \tag{11.71}$$

Thus, the GEKS indexes are not subject to chain drift within the window of T periods under consideration.

Recall definition (11.63) which defined the logarithm of the Törnqvist price index, $\ln P_T(i, j)$, that compared the prices of period j to the prices

[78] Eltetö and Köves (1964), Gini (1931) and Szulc (1964) independently derived the GEKS price indexes by an alternative route. Thus the name GEKS has the initials of all four primary authors of the method. Ivancic et al. (2009, 2011) suggested the use of the GEKS index in the time series context.

of period i. The GEKS methodology can be applied using $P_T(r, t)$ in place of the Fisher $P_F(r, t)$ as the basic bilateral index building block. Thus define the *period t GEKS Törnqvist price level*, p^t_{GEKST}, for $t = 1, \ldots, T$ as follows:

$$p^t_{\text{GEKST}} \equiv \left[\prod_{r=1}^{T} P_T(r, t)\right]^{1/T}. \tag{11.72}$$

The *GEKST price indexes* P^t_{GEKST} are obtained by normalizing the above price levels so that the period 1 index is equal to 1. Thus, we have the following definitions for P^t_{GEKST} for $t = 1, \ldots, T$:

$$P^t_{\text{GEKST}} \equiv \frac{p^t_{\text{GEKST}}}{p^1_{\text{GEKST}}}. \tag{11.73}$$

Since $P_T(r, t)$ approximates $P_F(r, t)$ to the second order around an equal price and quantity point, the P^t_{GEKST} will usually be quite close to the corresponding P^t_{GEKS} indexes.

It is possible to provide a very simple alternative approach to the derivation of the GEKS Törnqvist price indexes.[79] Define the *sample average sales share* for product n, $s_{\bullet n}$, and the *sample average log price* for product n, $\ln p_{\bullet n}$, as follows for $n = 1, \ldots, N$:

$$s_{\bullet n} \equiv \sum_{t=1}^{T} \frac{1}{T} s_{tn}; \tag{11.74}$$

$$\ln p_{\bullet n} \equiv \sum_{t=1}^{T} \frac{1}{T} \ln p_{tn}. \tag{11.75}$$

The logarithm of the *CCDI price level for period t*, $\ln p^t_{\text{CCDI}}$, is defined by comparing the prices of period t with the sample average prices using the bilateral Törnqvist formula; i.e., for $t = 1, \ldots, T$, we have the following definitions:

$$\ln p^t_{\text{CCDI}} \equiv \sum_{n=1}^{N} \frac{1}{2}(s_{tn} + s_{\bullet n})(\ln p_{tn} - \ln p_{\bullet n}). \tag{11.76}$$

[79] This approach is due to Inklaar and Diewert (2016). It is an adaptation of the distance function approach used by Caves et al. (1982) to the price index context.

The *CCDI price index for period t*, P^t_{CCDI} is defined as the following normalized CCDI price level for $t = 1, \ldots, T$:

$$P^t_{\text{CCDI}} \equiv \frac{p^t_{\text{CCDI}}}{p^1_{\text{CCDI}}}. \tag{11.77}$$

Using the above definitions, the logarithm of the CCDI price index for period t is equal to the following expressions for $t = 1, \ldots, T$:

$$\begin{aligned}
\ln P^t_{\text{CCDI}} &= \ln p^t_{\text{CCDI}} - \ln p^1_{\text{CCDI}} \\
&= \sum_{n=1}^{N} \frac{1}{2}(s_{tn} + s_{\bullet n})(\ln p_{tn} - \ln p_{\bullet n}) \\
&\quad - \sum_{n=1}^{N} \frac{1}{2}(s_{1n} + s_{\bullet n})(\ln p_{1n} - \ln p_{\bullet n}) \\
&= \ln P^t_T + \sum_{n=1}^{N} \frac{1}{2}(s_{tn} - s_{\bullet n})(\ln p_{1n} - \ln p_{\bullet n}) \\
&\quad - \sum_{n=1}^{N} \frac{1}{2}(s_{1n} - s_{\bullet n})(\ln p_{tn} - \ln p_{\bullet n}) \\
&= \ln P^t_{\text{GEKST}}
\end{aligned} \tag{11.78}$$

where the last equality follows by direct computation or by using the computations in Inklaar and Diewert (2016).[80] Thus the CCDI multilateral price indexes are equal to the GEKS Törnqvist multilateral indexes defined by (11.73). Define $s^\bullet \equiv [s_{\bullet 1}, \ldots, s_{\bullet N}]$ as the vector of sample average shares and $\ln p^\bullet \equiv [\ln p_{\bullet 1}, \ldots, \ln p_{\bullet N}]$ as the vector of sample average log prices. Then the last two terms on the right hand side of the penultimate equality in (11.78) can be written as $(1/2)N\text{cov}(s^t - s^\bullet, \ln p^1 - \ln p^\bullet) - (1/2)N\text{cov}(s^1 - s^\bullet, \ln p^t - \ln p^\bullet)$. If the fluctuations in shares and prices are not too violent, it is likely that both covariances are close to 0 and thus $\ln P^t_{\text{CCDI}} \approx \ln P^t_T$ for each t.[81]

[80] The second from last equality was derived in Diewert and Fox (2020).

[81] For Diewert's (2018) empirical example, the sample average of these two sets of covariance terms turned out to be 0 with variances equal to 0.00024 and 0.00036, respectively.

Thus under these circumstances, it is likely that $\ln P^t_{\text{CCDI}} \approx \ln P^t_T$ for each t. Moreover, under the assumptions of linear trends in log prices and linear trends in shares, assumptions (11.58) and (11.59), it was seen in the previous section that the period t bilateral Törnqvist price index, P^t_T, was equal to its chained counterpart for any t.[82] This result implies that $P^t_T = P^t_{\text{CCDI}} = P^t_{\text{GEKST}}$ for $t = 1, \ldots, T$ under the linear trends assumption. Thus we expect the period t multilateral index, $P^t_{\text{GEKST}} = P^t_{\text{CCDI}}$ to approximate the corresponding fixed base period t Törnqvist price index, P^t_T, provided that prices and quantities have smooth trends.

Since P^t_F approximates P^t_T, we expect that the following approximate equalities will hold under the smooth trends assumption for $t = 1, \ldots, T$:

$$P^t_F \approx P^t_T \approx P^t_{\text{GEKS}} \approx P^t_{\text{GEKST}} = P^t_{\text{CCDI}}. \tag{11.79}$$

The above indexes will be free from chain drift within the window of T periods[83]; i.e., if prices and quantities for any two periods in the sample are equal, then the price index will register the same value for these two periods.

Unit values taken over heterogeneous products are often used at the first stage of aggregation. In the following section, bias estimates for unit value price levels will be derived and in the subsequent section, quality adjusted unit value price levels will be studied.

Unit Value Price and Quantity Indexes

As was mentioned in section "Comparing CES Price Levels and Price Indexes", there was a preliminary aggregation over time problem that needed to be addressed; i.e., exactly how should the period t prices and quantities for commodity n, p^t_n and q^t_n, that are used in an index number formula be defined? During any time period t, there will typically be many transactions in a specific commodity n at a number of different

[82] See the discussion below (11.65) in the previous section. Note that the assumption of linear trends in shares is not consistent with the existence of new and disappearing products.

[83] See de Haan (2015) and Diewert and Fox (2020) for discussions of the problems associated with linking the results from one rolling window multilateral comparison to a subsequent window of observations. Empirically, there does not appear to be much chain drift when the indexes generated by subsequent windows are linked.

prices. Hence, there is a need to provide a more precise definition for the "average" or "representative" price for commodity n in period t, p_n^t. Starting with Drobisch (1871), many measurement economists and statisticians advocated the use of the *unit value* (total value transacted divided by total quantity) as the appropriate price p_n^t for commodity n and the total quantity transacted during period t as the appropriate quantity, q_n^t, e.g., see Walsh (1901, 96; 1921a, 88), Fisher (1922, 318) and Davies (1924, 183; 1932, 59). If it is desirable to have q_n^t be equal to the total quantity of commodity n transacted during period t and also desirable to have the product of the price p_n^t times quantity q_n^t to be equal the value of period t transactions in commodity n, then one is *forced* to define the aggregate period t price for commodity n, p_n^t, to be the total value transacted during the period divided by the total quantity transacted, which is the unit value for commodity n.[84]

There is general agreement that a unit value price is an appropriate price concept to be used in an index number formula if the transactions refer to a narrowly defined homogeneous commodity. Our task in this section is to look at the properties of a unit value price index when aggregating over commodities that are *not* completely homogeneous. We will also look at the properties of the companion unit value quantity index in this section.

The period t *unit value price level*, p_{UV}^t, and the corresponding period t *unit value price index* which compares the price level in period t to that of period 1, P_{UV}^t, are defined as follows for $t = 1, \ldots, T$:

$$p_{\text{UV}}^t \equiv \frac{p^t \cdot q^t}{1_N \cdot q^t}; \tag{11.80}$$

$$\begin{aligned} P_{\text{UV}}^t &\equiv \frac{p_{\text{UV}}^t}{p_{\text{UV}}^1} \\ &= \frac{\left[p^t \cdot q^t / 1_N \cdot q^t\right]}{\left[p^1 \cdot q^1 / 1_N \cdot q^1\right]} \\ &= \frac{\left[p^t \cdot q^t / p^1 \cdot q^1\right]}{Q_{\text{UV}}^t} \end{aligned} \tag{11.81}$$

[84] For additional discussion on unit value price indexes, see Balk (2008, 72–74), Diewert and von der Lippe (2010), Silver (2010, 2011) and de Haan and Krsinich (2018).

where the period t *unit value quantity index*, Q^t_{UV}, is defined as follows for $t = 1, \ldots, T$:

$$Q^t_{UV} \equiv \frac{1_N \cdot q^t}{1_N \cdot q^1}. \tag{11.82}$$

It can be seen that the unit value price index satisfies Walsh's multiperiod identity test and thus P^t_{UV} is free from chain drift.

However, there is a big problem in using the unit value price index when the commodities in scope are not homogeneous: *the unit value price index is not invariant to changes in the units of measurement of the individual products in the aggregate*.

We will look at the relationship of the *unit value quantity indexes*, Q^t_{UV}, with the corresponding *Laspeyres*, *Paasche* and *Fisher fixed base quantity indexes*, Q^t_L, Q^t_P and Q^t_F, defined below for $t = 1, \ldots, T$:

$$Q^t_L \equiv \frac{p^1 q^t}{p^1 q^1} = \sum_{n=1}^{N} s_{1n}\left(\frac{q_{tn}}{q_{1n}}\right); \tag{11.83}$$

$$Q^t_P \equiv \frac{p^t q^t}{p^t q^1} = \left[\sum_{n=1}^{N} s_{tn}\left(\frac{q_{tn}}{q_{1n}}\right)^{-1}\right]^{-1}; \tag{11.84}$$

$$Q^t_F \equiv [Q^t_L Q^t_P]^{1/2}. \tag{11.85}$$

For the second set of equations in (11.83), we require that $q_{1n} > 0$ for all n and for the second set of equations in (11.84), we require that all $q_{tn} > 0$. Recall that the period t *sales* or *expenditure share* vector $s^t \equiv [s_{t1}, \ldots, s_{tN}]$ was defined at the beginning of section "Comparing CES Price Levels and Price Indexes". The period t *quantity* share vector $S^t \equiv [S_{t1}, \ldots, S_{tN}]$ was also defined in section "Comparing CES Price Levels and Price Indexes" as follows for $t = 1, \ldots, T$:

$$S^t \equiv \frac{q^t}{1_N \cdot q^t}. \tag{11.86}$$

Below, we will make use of the following identities (11.87), which hold for $t = 1, \ldots, T$:

$$\sum_{n=1}^{N} [p^t_{UV} - p_{tn}]q_{tn} = \sum_{n=1}^{N}\left(\frac{p^t \cdot q^t}{1_N \cdot q^t} - p_{tn}\right)q_{tn} \tag{11.87}$$

using definitions (11.80)

$$= \left(\frac{p^t \cdot q^t}{1_N \cdot q^t}\right) 1_N \cdot q^t - p^t \cdot q^t$$
$$= 0.$$

The following relationships between Q^t_{UV} and Q^t_L hold for $t = 1, \ldots, T$:

$$Q^t_{UV} - Q^t_L = \frac{1_N \cdot q^t}{1_N \cdot q^1} - \frac{p^1 \cdot q^t}{p^1 \cdot q^1} \tag{11.88}$$

using (11.82) and (11.83)

$$= \sum_{n=1}^{N} S_{1n}\left(\frac{q_{tn}}{q_{1n}}\right) - \sum_{n=1}^{N} s_{1n}\left(\frac{q_{tn}}{q_{1n}}\right)$$

using (11.86) and (11.83)

$$= \sum_{n=1}^{N} [S_{1n} - s_{1n}]\left(\frac{q_{tn}}{q_{1n}}\right)$$
$$= N\text{cov}\left(S^1 - s^1,\ q^t / q^1\right)$$

where the vector of period t to period 1 relative quantities is defined as $q^t / q^1 \equiv \left[q_{t1}/q_{11}, q_{t2}/q_{12}, \ldots, q_{tN}/q_{1N}\right]$. As usual, there are three special cases of (11.88) which will imply that $Q^t_{\text{UV}} = Q^t_L$. (i) $S^1 = s^1$ so that the vector of period 1 real quantity shares S^1 is equal to the period 1 sales share vector s^1. This condition is equivalent to $p^1 = \lambda_1 1_N$ so that all period 1 prices are equal.[85] (ii) $q^t = \lambda_t q^1$ for $t = 2, 3, \ldots, T$ so that quantities vary in strict proportion over time. (iii) $\text{cov}\left(S^1 - s^1, q^t / q^1\right) = 0$.[86]

[85] Consider the case where $p^1 = \lambda 1_N$. Units of measurement for the N commodities can always be chosen so that all prices are equal in period 1. Then $Q^t_{\text{UV}} = Q^t_L$ and hence $P^t_{\text{UV}} = P^t_P$ where P^t_{UV} is defined by (11.81) and P^t_P is the fixed base Paasche price index defined by (11.34). Thus, for this particular choice for units of measurement, the unit value price index P^t_{UV} is equal to a fixed base Paasche price index which will typically have a downward bias relative to a superlative index.

[86] For similar bias formulae, see Balk (2008, 73–74) and Diewert and von der Lippe (2010).

There are two problems with the above bias formula: (i) it is difficult to form a judgment on the sign of the covariance $\mathrm{cov}\big(S^1 - s^1, q^t/q^1\big)$ and (ii) the decomposition given by (11.88) requires that all components of the period 1 quantity vector be positive.[87] It would be useful to have a decomposition that allowed some quantities (and sales shares) to be equal to 0. Consider the following alternative decomposition to (11.88) for $t = 1, \ldots, T$:

$$Q^t_{\mathrm{UV}} - Q^t_L = \left[\frac{1_N \cdot q^t}{1_N \cdot q^1}\right] - \left[\frac{p^1 \cdot q^t}{p^1 \cdot q^1}\right] \tag{11.89}$$

using (11.82) and (11.83)

$$= \sum_{n=1}^{N} \left[\left(\frac{q_{tn}}{1_N \cdot q^1}\right) - \left(\frac{p_{1n} q_{tn}}{p^1 \cdot q^1}\right)\right]$$

$$= \sum_{n=1}^{N} \left[\left(\frac{1}{1_N \cdot q^1}\right) - \left(\frac{p_{1n}}{p^1 \cdot q^1}\right)\right] q_{tn}$$

$$= \sum_{n=1}^{N} \left[\left(\frac{p^1 \cdot q^1}{1_N \cdot q^1}\right) - p_{1n}\right]\left[\frac{q_{tn}}{p^1 \cdot q^1}\right]$$

$$= \sum_{n=1}^{N} \left[p^1_{UV} - p_{1n}\right]\left[\frac{q_{tn}}{p^1 \cdot q^1}\right]$$

using (11.80) for $t = 1$

$$= \sum_{n=1}^{N} \frac{\left[p^1_{UV} - p_{1n}\right]\left[q_{tn} - q_{1n} Q^t_{UV}\right]}{p^1 \cdot q^1}$$

using (11.87) for $t = 1$

$$= Q^t_{\mathrm{UV}} \sum_{n=1}^{N} \frac{\left(p^1_{\mathrm{UV}} - p_{1n}\right)\left[\left(q_{tn}/Q^t_{\mathrm{UV}}\right) - q_{1n}\right]}{p^1 \cdot q^1}$$

[87] We are assuming that all prices are positive in all periods (so if there are missing prices they must be replaced by positive imputed prices) but we are not assuming that all quantities (and expenditure shares) are positive.

$$= Q_{UV}^t \sum_{n=1}^{N} s_{1n} \left[\frac{p_{UV}^1}{p_{1n}} - 1 \right] \left[\frac{q_{tn}}{q_{1n} Q_{UV}^t} - 1 \right] \quad \text{if } q_{1n} > 0 \text{ for all } n$$
$$= Q_{UV}^t \varepsilon_L^t$$

where the *period t error term* ε_L^t is defined for $t = 1, \ldots, T$ as[88]:

$$\varepsilon_L^t \equiv \sum_{n=1}^{N} \frac{\left[p_{UV}^1 - p_{1n}\right]\left[\left(\frac{q_{tn}}{Q_{UV}^t}\right) - q_{1n}\right]}{p^1 \cdot q^1}. \tag{11.90}$$

If $q_{1n} > 0$ for $n = 1, \ldots, N$, then ε_L^t is equal to $\sum_{n=1}^{N} s_{1n}\left[\left(p_{UV}^1/p_{1n}\right) - 1\right]\left[\left(q_{tn}/q_{1n}Q_{UV}^t\right) - 1\right]$.

Note that the terms on the right hand side of (11.90) can be interpreted as $\left(N/p^1 \cdot q^1\right)$ times the covariance $\text{cov}\left(p_{UV}^1 1_N - p^1, q^t - Q_{UV}^t q^1\right)$ since $1_N \cdot \left(q^t - Q_{UV}^t q^1\right) = 0$. If the products are substitutes, it is likely that this covariance is *negative*, since if p_{1n} is unusually low, we would expect that it would be less than the period 1 unit value price level p_{UV}^1 so that $p_{UV}^1 - p_{1n} > 0$. Furthermore, if p_{1n} is unusually low, then we would expect that the corresponding q_{1n} is unusually high, and thus it is likely that q_{1n} is greater than q_{tn}/Q_{UV}^t and so $q_{tn} - q_{1n}Q_{UV}^t < 0$. Thus the N terms in the covariance will tend to be negative provided that there is some degree of substitutability between the products.[89] Looking at formula (11.90) for ε_L^t, it can be seen that all terms on the right hand side of (11.90) do not depend on t, except for the N period t deflated product quantity terms, q_{tn}/Q_{UV}^t for $n = 1, \ldots, N$. Hence, if there is a great deal of variation in the period t quantities q_{tn}, then $\left(q_{tn}/Q_{UV}^t\right) - q_{1n}$ could be positive or negative and thus the tendency for ε_L^t to be negative will be a weak one. Thus our expectation is that the error term ε_L^t is likely to be negative and hence $Q_{UV}^t < Q_L^t$ for $t \geq 2$ but this expectation is a weak one.

[88] Note that this error term is homogeneous of degree 0 in the components of p^1, q^1 and q^t. Hence, it is invariant to proportional changes in the components of these vectors.

[89] The results in previous sections looked at responses of product *shares* to changes in prices and with data that are consistent with CES preferences, the results depended on whether the elasticity of substitution was greater or less than unity. In the present section, the results depend on whether the elasticity of substitution is equal to 0 or greater than 0; i.e., it is the response of *quantities* (rather than *shares*) to lower prices that matters.

It should be noted that P^t_{UV} and Q^t_{UV} *do not depend on the estimated reservation prices* for the missing products; i.e., the definitions of P^t_{UV} and Q^t_{UV} zero out the estimated reservation prices.

As usual, there are 3 special cases of (11.89) that will imply that $Q^t_{UV} = Q^t_L$: (i) $p^1 = \lambda_1 1_N$ so that all period 1 prices are equal; (ii) $q^t = \lambda_t q^1$ for $t = 2, 3, \ldots, T$ so that quantities vary in strict proportion over time; (iii) $\mathrm{cov}(p^1_{UV} 1_N - p^1, q^t - Q^t_{UV} q^1) = 0$. These conditions are equivalent to our earlier conditions listed below (11.88).

If we divide both sides of equation t in Eq. (11.89) by Q^t_{UV}, we obtain the following system of identities for $t = 1, \ldots, T$:

$$\frac{Q^t_L}{Q^t_{UV}} = 1 - \varepsilon^t_L \tag{11.91}$$

where we expect ε^t_L to be a small negative number in the elementary index context.

The identities in (11.89) and (11.91) are valid if we interchange prices and quantities. The quantity counterparts to p^t_{UV} and P^t_{UV} defined by (11.80) and (11.81) are the period t *Dutot quantity level* q^t_D and *quantity index* Q^t_D[90] defined as $q^t_D \equiv p^t \cdot q^t / 1_N \cdot p^t = \alpha^t \cdot q^t$ (where $\alpha^t \equiv p^t / 1_N \cdot p^t$ is a vector of period t price weights for q^t) and $Q^t_D \equiv q^t_{UV} / q^1_{UV} = [p^t \cdot q^t / p^1 \cdot q^1] / P^t_D$ where we redefine the period t Dutot price level as $p^t_D \equiv 1_N \cdot p^t$ and the period t Dutot price index as $P^t_D \equiv p^t_D / p^1_D = 1_N \cdot p^t / 1_N \cdot p^1$ which coincides with our earlier definition (11.10) for P^t_D. Using these definitions and interchanging prices and quantities, Eq. (11.91) become the following equations for $t = 1, \ldots, T$:

$$\frac{P^t_L}{P^t_D} = 1 - \varepsilon^{t*}_L \tag{11.92}$$

where the period t error term ε^{t*}_L is defined for $t = 1, \ldots, T$ as:

$$\varepsilon^{t*}_L \equiv \sum_{n=1}^{N} \frac{[q^1_D - q_{1n}][(p_{tn} / P^t_D) - p_{1n}]}{p^1 \cdot q^1}. \tag{11.93}$$

If p_{1n} is unusually low, then it is likely that it will be less than p_{tn} / P^t_D and it is also likely that q_{1n} will be unusually high and hence greater than

[90] Balk (2008, 7) called Q^t_{UV} a Dutot-type quantity index.

the average period 1 Dutot quantity level, q_D^1. Thus the N terms in the definition of ε_L^{t*} will tend to be negative and thus $1 - \varepsilon_L^{t*}$ will tend to be greater than 1. Thus there will be a tendency for $P_D^t < P_L^t$ for $t \geq 2$ but again, this expectation is a weak one if there are large fluctuations in the deflated period t prices, p_{tn} / P_D^t for $n = 1, \ldots, N$.

It can be verified that the following identities hold for the period t Laspeyres, Paasche and unit value price and quantity indexes for $t = 1, \ldots, T$:

$$\frac{p^t \cdot q^t}{p^1 \cdot q^1} = P_{UV}^t Q_{UV}^t = P_P^t Q_L^t = P_L^t Q_P^t. \tag{11.94}$$

Equation (11.94) imply the following identities for $t = 1, \ldots, T$:

$$\begin{aligned} \frac{P_{\mathrm{UV}}^t}{P_P^t} &= \frac{Q_L^t}{Q_{\mathrm{UV}}^t} \\ &= 1 - \varepsilon_L^t \end{aligned} \tag{11.95}$$

where the last set of equations follow from Eq. (11.91). Thus, we expect that $P_{\mathrm{UV}}^t > P_P^t$ for $t = 2, 3, \ldots, T$ if the products are substitutes and ε_L^t is negative.[91]

We now turn our attention to developing an exact relationship between Q_{UV}^t and the Paasche quantity index Q_P^t. Using definitions (11.82) and (11.84), we have for $t = 1, \ldots, T$:

$$\left(Q_{\mathrm{UV}}^t\right)^{-1} - \left(Q_P^t\right)^{-1} = \left[\frac{1_N \cdot q^1}{1_N \cdot q^t}\right] - \left[\frac{p^t \cdot q^1}{p^t \cdot q^t}\right] \tag{11.96}$$

using (11.82) and (11.84)

$$\begin{aligned} &= \sum_{n=1}^{N} [S_{tn} - s_{tn}] \left[\frac{q_{1n}}{q_{tn}}\right] \\ &= N \mathrm{cov}\left(S^t - s^t, \frac{q^1}{q^t}\right) \end{aligned}$$

[91] As was discussed earlier, if all prices are equal in the base period, then $\varepsilon_L^t = 0$ and $P_{\mathrm{UV}}^t / P_P^t = Q_L^t / Q_{\mathrm{UV}}^t = 0$.

where the second set of equalities in (11.96) follows using (11.88) and (11.86), assuming that $q_{tn} > 0$ for $n = 1, \ldots, N$..

As usual, there are three special cases of (11.96) that will imply that $Q^t_{\text{UV}} = Q^t_P$: (i) $S^t = s^t$ so that the vector of period t real quantity shares S^t is equal to the period t sales share vector s^t. This condition is equivalent to $p^t = \lambda_t 1_N$ which implies that all period t prices are equal.[92] (ii) $q^t = \lambda_t q^1$ for $t = 2, 3, \ldots, T$ so that quantities vary in strict proportion over time. (iii) $N\text{cov}\left(S^t - s^t, q^1/q^t\right) = 0$.

Again, there are two problems with the above bias formula: (i) it is difficult to form a judgment on the sign of the covariance $N\text{cov}\left(S^t - s^t, q^1/q^t\right)$ and (ii) the decomposition given by (11.96) requires that all components of the period t quantity vector be positive. We will proceed to develop a decomposition that does not require the positivity of q^t. The following exact decomposition holds for $t = 1, \ldots, T$:

$$\begin{aligned}
[Q^t_{\text{UV}}]^{-1} - [Q^t_P]^{-1} &= \frac{1_N \cdot q^1}{1_N \cdot q^t} - \frac{p^t \cdot q^1}{p^t \cdot q^t} \\
&= \sum_{n=1}^{N} \left[\frac{q_{1n}}{1_N \cdot q^t} - \frac{p_{tn} q_{1n}}{p^t \cdot q^t} \right] \\
&= \sum_{n=1}^{N} \left[\left(\frac{1}{1_N \cdot q^t} \right) - \left(\frac{p_{tn}}{p^t \cdot q^t} \right) \right] q_{1n} \\
&= \sum_{n=1}^{N} \left[\left(\frac{p^t \cdot q^t}{1_N \cdot q^t} \right) - p_{tn} \right] \left[\frac{q_{1n}}{p^t \cdot q^t} \right] \\
&= \sum_{n=1}^{N} [p^t_{\text{UV}} - p_{tn}] \left[\frac{q_{1n}}{p^t \cdot q^t} \right] \qquad (11.97)
\end{aligned}$$

[92] If $p^t = \lambda 1_N$, so that all prices are equal in period t, then it can be shown directly that $P^t_{\text{UV}} = P^t_L$. Thus for the particular choice for units of measurement that makes all prices equal in period t, the unit value price index P^t_{UV} is equal to a fixed base Laspeyres price index which will typically have an upward bias relative to a superlative index.

using (11.80)

$$= \sum_{n=1}^{N} \frac{[p_{\mathrm{UV}}^t - p_{tn}][q_{1n} - (q_{tn}/Q_{\mathrm{UV}}^t)]}{p^t \cdot q^t}$$

using (11.87)

$$= (Q_{\mathrm{UV}}^t)^{-1} \sum_{n=1}^{N} \frac{[p_{\mathrm{UV}}^t - p_{tn}][q_{1n} Q_{\mathrm{UV}}^t - q_{tn}]}{p^t \cdot q^t}$$

$$= (Q_{\mathrm{UV}}^t)^{-1} \sum_{n=1}^{N} s_{tn} \left[\frac{p_{\mathrm{UV}}^t}{p_{tn}} - 1\right]\left[\frac{q_{1n} Q_{\mathrm{UV}}^t}{q_{tn}} - 1\right] \quad \text{if } q_{tn} > 0 \text{ for all } n$$

$$= (Q_{\mathrm{UV}}^t)^{-1} \varepsilon_P^t$$

where the period t error term ε_P^t is defined as follows for $t = 1, \ldots, T$[93]:

$$\varepsilon_P^t \equiv \sum_{n=1}^{N} \frac{[p_{\mathrm{UV}}^t - p_{tn}][q_{1n} Q_{\mathrm{UV}}^t - q_{tn}]}{p^t \cdot q^t}. \tag{11.98}$$

If $q_{tn} > 0$ 0 for $n = 1, \ldots, N$, then ε_P^t is equal to $\sum_{n=1}^{N} s_{tn}[(p_{\mathrm{UV}}^t/p_{tn}) - 1][(q_{1n} Q_{\mathrm{UV}}^t/q_{tn}) - 1]$.

Note that the terms on the right hand side of (11.97) can be interpreted as $(N/p^t \cdot q^t)$ times the covariance $\mathrm{cov}\left(p_{\mathrm{UV}}^t 1_N - p^t, q^1 - (Q_{\mathrm{UV}}^t)^{-1} q^t\right)$ since $1_N \cdot \left(q^1 - (Q_{UV}^t)^{-1} q^t\right) = 0$. If the products are substitutable, it is likely that this covariance is *negative*, since if p_{tn} is unusually low, we would expect that it would be less than the period t unit value price p_{UV}^t so that $p_{\mathrm{UV}}^t - p_{tn} > 0$. If p_{tn} is unusually

[93] Note that this error term is homogeneous of degree 0 in the components of p^t, q^1 and q^t. Thus for $\lambda > 0$, we have $\varepsilon_P\left(p^t, q^1, q^t\right) = \varepsilon_P\left(\lambda p^t, q^1, q^t\right) = \varepsilon_P\left(p^t, \lambda q^1, q^t\right) = \varepsilon_P\left(p^t, q^1, \lambda q^t\right)$. Note also that ε_P^t is well defined if some quantities are equal to 0 and ε_P^t does depend on the reservation prices p_{tn} for products n that are not present in period t. If product n is missing in period t, then it is likely that the reservation price p_{tn} is greater than the unit value price level for period t, p_{UV}^t, and since $q_{tn} = 0$, it can be seen that the nth term on the right hand side of (11.98) will be negative; i.e., the greater the number of missing products in period t, the greater is the likelihood that ε_P^t is negative.

low, then we also expect that the corresponding q_{tn} is unusually high, and thus it is likely that q_{tn} is greater than $q_{1n}Q^t_{\text{UV}}$ and so $q_{1n}Q^t_{\text{UV}} - q_{tn} < 0$. Thus the N terms in the covariance will tend to be negative. Thus our expectation is that the error term $\varepsilon^t_P < 0$ and $\left[Q^t_{\text{UV}}\right]^{-1} < \left[Q^t_P\right]^{-1}$ or $Q^t_{\text{UV}} > Q^t_P$ for $t \geq 2$.[94]

There are three special cases of (11.97) that will imply that $Q^t_{\text{UV}} = Q^t_P$: (i) $p^t = \lambda_t 1_N$ so that all period t prices are equal; (ii) $q^t = \lambda_t q^1$ for $t = 2, 3, \ldots, T$ so that quantities vary in strict proportion over time; (iii) $\text{cov}\left(p^t_{\text{UV}}1_N - p^t, q^1 - \left[Q^t_{\text{UV}}\right]^{-1}q^t\right) = 0$. These conditions are equivalent to our earlier conditions listed below (11.96).

If we divide both sides of equation t in Eq. (11.97) by $\left[Q^t_{\text{UV}}\right]^{-1}$, we obtain the following system of identities for $t = 1, \ldots, T$:

$$\frac{Q^t_P}{Q^t_{\text{UV}}} = \left[1 - \varepsilon^t_P\right]^{-1} \tag{11.99}$$

where we expect ε^t_P to be a small negative number if the products are substitutable. Thus, we expect $Q^t_P < Q^t_{\text{UV}} < Q^t_L$ for $t = 2, 3, \ldots, T$.

Equations (11.97) and (11.99) are valid if we interchange prices and quantities. Using the definitions for the Dutot price and quantity levels and indexes t and interchanging prices and quantities, Eq. (11.99) become $P^t_P / P^t_D = \left[1 - \varepsilon^{t*}_P\right]^{-1}$ where $\varepsilon^{t*}_P \equiv \sum_{n=1}^N \left[q^t_D - q_{tn}\right]\left[\left(p_{1n}P^t_D\right) - p_{tn}\right] / p^t \cdot q^t$ for $t = 1, \ldots, T$. If p_{tn} is unusually low, then it is likely that it will be less than p_{tn} / P^t_D and it is also likely that q_{tn} will be unusually high and hence greater than the average period t Dutot quantity level q^t_D. Thus the N terms in the definition of ε^{t*}_P will tend to be negative and hence a tendency for $\left[1 - \varepsilon^{t*}_P\right]^{-1}$ to be less than 1. Thus, there will be a tendency for $P^t_P < P^t_D$ for $t \geq 2$.

Equation (11.94) imply the following identities for $t = 1, \ldots, T$:

$$\frac{P^t_{\text{UV}}}{P^t_L} = \frac{Q^t_P}{Q^t_{\text{UV}}} = \left[1 - \varepsilon^t_P\right]^{-1} \tag{11.100}$$

[94] Our expectation that ε^t_P is negative is more strongly held than our expectation that ε^t_L is negative.

where the last set of equations follow from Eq. (11.99). Thus, we expect that $P_P^t < P_{UV}^t < P_L^t$ for $t = 2, 3, \ldots, T$ if the products are substitutes.[95]

Equations (11.95) and (11.100) develop exact relationships for the unit value price index P_{UV}^t with the corresponding fixed base Laspeyres and Paasche price indexes, P_L^t and P_P^t. Taking the square root of the product of these two sets of equations leads to the following exact relationships between the fixed base Fisher price index, P_F^t, and its unit value counterpart period t index, P_{UV}^t, for $t = 1, \ldots, T$:

$$P_{UV}^t = P_F^t \left\{ \frac{1 - \varepsilon_L^t}{1 - \varepsilon_P^t} \right\}^{1/2} \tag{11.101}$$

where ε_L^t and ε_P^t are defined by (11.90) and (11.98). If there are no strong (divergent) trends in prices and quantities, then it is likely that ε_L^t is approximately equal to ε_P^t and hence under these conditions, it is likely that $P_{UV}^t \approx P_F^t$; i.e., the unit value price index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, with diverging trends in prices and quantities (in opposite directions), we would expect the error term ε_P^t defined by (11.98) to be more negative than the error term ε_L^t defined by (11.90) and thus under these conditions, we expect the unit value price index P_{UV}^t to have a *downward bias* relative to its Fisher price index counterpart P_F^t.[96]

However, if there are missing products in period 1 so that some q_{1n} are equal to 0 and the corresponding imputed prices p_{1n} are greater than the unit value price for observation 1, p_{UV}^1, then the nth term in the sum of terms on the right hand side of (11.90) can become negative and large in magnitude, which can make ε_L^t defined by (11.90) much more negative

[95] If $p^t = \lambda 1_N$, then $\varepsilon_P^t = 0$, $P_{UV}^t = P_L^t$ and $Q_{UV}^t = Q_P^t$. Thus if prices in period t are all equal, the period t fixed base unit value index will equal the fixed base Laspeyres price index. Thus, the unit value index will tend to have an upward bias relative to a superlative index in this equal period t prices case.

[96] The Dutot price index counterparts to the exact relations (11.101) are $P_F^t = P_D^t\left\{\left(1 - \varepsilon_L^{t*}\right)/\left(1 - \varepsilon_P^{t*}\right)\right\}^{1/2}$ for $t = 1, \ldots, T$. Thus with diverging trends in prices and quantities (in opposite directions), we would expect the error term ε_P^{t*} to be more negative than the error term ε_L^{t*} and hence we would expect $P_D^t > P_F^t$ for $t \geq 2$. Note that the Dutot price index can be interpreted as a *fixed basket price index* where the basket is proportional to a vector of ones. Thus, with divergent trends in prices and quantities in opposite directions, we would expect the Dutot index to exhibit substitution bias and hence we would expect $P_D^t > P_F^t$ for $t \geq 2$.

than ε_P^t, which in turn means that P_{UV}^t will be greater than unit value price index P_F^t using (11.101) above. Thus, under these circumstances, the unit value price index P_{UV}^t will have an *upward bias* relative to its Fisher price index counterpart P_F^t.

It is possible that unit value price indexes can approximate their Fisher counterparts to some degree in some circumstances but these approximations are not likely to be very accurate. If the products are somewhat heterogeneous and there are some divergent trends in price and quantities, then the approximations are likely to be poor.[97] They are also likely to be poor if there is substantial product turnover.

Quality Adjusted Unit Value Price and Quantity Indexes

In the previous section, the period t nit value quantity *level* was defined by $q_{UV}^t \equiv 1_N \cdot q^t = \sum_{n=1}^{N} q_{tn}$ for $t = 1, \ldots, T$. The corresponding period t unit value quantity *index* was defined by (11.82) for $t = 1, \ldots, T$; i.e., $Q_{UV}^t \equiv 1_N \cdot q^t / 1_N \cdot q^1$. In the present section, we will consider *quality adjusted unit value quantity levels*, $q_{UV\alpha}^t$ and the corresponding *quality adjusted unit value quantity indexes*, $Q_{UV\alpha}^t$, defined as follows for $t = 1, \ldots, T$:

$$q_{UV\alpha}^t \equiv \alpha \cdot q^t; \tag{11.102}$$

$$Q_{UV\alpha}^t \equiv \frac{q_{UV\alpha}^t}{q_{UV\alpha}^1} = \frac{\alpha \cdot q^t}{\alpha \cdot q^1} \tag{11.103}$$

where $\alpha \equiv [\alpha_1, \ldots, \alpha_N]$ is a vector of positive *quality adjustment factors*. Note that if consumers value their purchases of the N products according to the linear utility function $f(q) \equiv \alpha \cdot q$, then the period t quality adjusted aggregate quantity level $q_{UV\alpha}^t = \alpha \cdot q^t$ can be interpreted as the aggregate (sub) *utility* of consumers of the N products. Note that this utility function is linear and thus the products are perfect substitutes, after adjusting for the relative quality of the products. The bigger α_N is,

[97] The problem with unit value price indexes is that they correspond to an additive quantity level. If one takes the economic approach to index number theory, then an additive quantity level corresponds to a linear utility function which implies an infinite elasticity of substitution between products, which is too high in general.

the more consumers will value a unit of product n over other products. The period t *quality adjusted unit value price level* and *price index*, $p^t_{\text{UV}\alpha}$ and $P^t_{\text{UV}\alpha}$, are defined as follows for $t = 1, \ldots, T$:

$$p^t_{\text{UV}\alpha} \equiv \frac{p^t \cdot q^t}{q^t_{\text{UV}\alpha}} = \frac{p^t \cdot q^t}{\alpha \cdot q^t}; \tag{11.104}$$

$$P^t_{\text{UV}\alpha} \equiv \frac{p^t_{\text{UV}\alpha}}{p^1_{\text{UV}\alpha}} = \frac{\left[p^t \cdot q^t / p^1 \cdot q^1\right]}{Q^t_{\text{UV}\alpha}}. \tag{11.105}$$

It is easy to check that the quality adjusted unit value price index satisfies Walsh's multiperiod identity test and thus is free from chain drift.[98] Note that the $P^t_{UV\alpha}$ and $Q^t_{\text{UV}\alpha}$ *do not depend on any estimated reservation prices*; i.e., the definitions of $P^t_{\text{UV}\alpha}$ and $Q^t_{\text{UV}\alpha}$ zero out any reservation prices that are applied to missing products.

Quality adjusted unit value price indexes are consistent with the economic approach to index number theory. If consumers of the N products under consideration all have linear utility functions of the form $f(q) \equiv \alpha q = \sum_{n=1}^{N} \alpha_n q_n$, then $Q^t_{\text{UV}\alpha}$ defined by (11.103) accurately represents real welfare growth going from period 1 to t and $P^t_{UV\alpha}$ defined by (11.105) represents consumer inflation over this period. It does not matter if there are new or disappearing products over this period; aggregate welfare or utility for period t is well defined as $\sum_{n=1}^{N} \alpha_n q_{tn}$ even if some q_{tn} are equal to 0. If $q_{tn=0}$, then the contribution of product n to utility in period t is $\alpha_n q_n = 0$. Furthermore, the quality adjusted unit value price and quantity indexes are invariant to changes in the units of measurement if we make the convention that if the units of measurement of q_n are changed to $\lambda_n q_n$ for some positive constant λ_n, then the corresponding α_n is changed to α_n / λ_n.[99] Note that regular unit value price indexes are not invariant to changes in the units of measurement.

From the viewpoint of the economic approach to index number theory, the problem with quality adjusted unit value price and quantity indexes

[98] The term "quality adjusted unit value price index" was introduced by Dalén (2001). Its properties were further studied by de Haan (2004b, 2010) and de Haan and Krsinich (2018). Von Auer (2014) considered a wide variety of choices for the weight vector  (including $\alpha = p^1$ and $\alpha = p^t$) and he looked at the axiomatic properties of the resulting indexes.

[99] Some methods for estimating the α_n are suggested in Diewert and Feenstra (2017) and Diewert (2021c).

is that the underlying linear utility function assumes that the N products under consideration are perfect substitutes after quality adjustment. Linear preferences are a special case of Constant Elasticity of Substitution preferences and the elasticity of substitution for linear preferences is equal to plus infinity. Empirical estimates for the elasticity of substitution are far less than plus infinity.[100]

We will start out by comparing $Q^t_{\mathrm{UV}\alpha}$ to the corresponding Laspeyres, Paasche and Fisher period t quantity indexes, Q^t_L, Q^t_P and Q^t_F. The algebra in this section follows the algebra in the preceding section. Thus the counterparts to the identities (11.87) in the previous section are the following identities for $t = 1, \ldots, T$:

$$\sum_{n=1}^{N} [\alpha_n p^t_{\mathrm{UV}\alpha} - p_{tn}]q_{tn} = \sum_{n=1}^{N} \left[\alpha_n \left(\frac{p^t \cdot q^t}{\alpha \cdot q^t}\right) - p_{tn}\right] q_{tn} \tag{11.106}$$

using definitions (11.104)

$$= \left(\frac{p^t \cdot q^t}{\alpha \cdot q^t}\right) \alpha \cdot q^t - p^t \cdot q^t$$
$$= 0.$$

The difference between the quality adjusted unit value quantity index for period t, $Q^t_{\mathrm{UV}\alpha}$ and the Laspeyres quantity index for period t, Q^t_L, can be written as follows for $t = 1, \ldots, T$:

$$Q^t_{\mathrm{UV}\alpha} - Q^t_L = \left[\frac{\alpha \cdot q^t}{\alpha \cdot q^1}\right] - \left[\frac{p^1 \cdot q^t}{p^1 \cdot q^1}\right] \tag{11.107}$$

[100] Quality adjusted unit value price and quantity levels are also consistent with Leontief (no substitition) preferences. In this case, the dual unit cost function is equal to $c(p) \equiv \sum_{n=1}^{N} \beta_n p_n$ where the β_n are positive preference parameters. The period t quantity vector that is consistent with these preferences is $q^t = u_t \beta$ for $t = 1, \ldots, T$ where $\beta \equiv [\beta_1, \ldots, \beta_N]$ and u_t is the period t utility level. Thus, the quantity vectors q^t will vary in strict proportion over time. This model of consumer behavior is inconsistent with situations where there are new and disappearing products over the T periods. Moreover, empirically, quantity vectors do not vary in a proportional manner over time.

using (11.83) and (11.103)

$$= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n q_{tn}}{\alpha \cdot q^1} \right) - \left(\frac{p_{1n} q_{tn}}{p^1 \cdot q^1} \right) \right]$$

$$= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n}{\alpha \cdot q^1} \right) - \left(\frac{p_{1n}}{p^1 \cdot q^1} \right) \right] q_{tn}$$

$$= \sum_{n=1}^{N} \left[\frac{\alpha_n p^1 \cdot q^1}{\alpha \cdot q^1} - p_{1n} \right] \left[\frac{q_{tn}}{p^1 \cdot q^1} \right]$$

$$= \sum_{n=1}^{N} \left[\alpha_n p^1_{\text{UV}\alpha} - p_{1n} \right] \left[\frac{q_{tn}}{p^1 \cdot q^1} \right]$$

using (11.104) for $t = 1$

$$= \sum_{n=1}^{N} \frac{[\alpha_n p^1_{\text{UV}\alpha} - p_{1n}][q_{tn} - q_{1n} Q^t_{\text{UV}\alpha}]}{p^1 \cdot q^1}$$

using (11.106) for $t = 1$

$$= Q^t_{\text{UV}\alpha} \sum_{n=1}^{N} \alpha_n \frac{[p^1_{\text{UV}\alpha} - (p_{1n}/\alpha_n)][(q_{tn}/Q^t_{\text{UV}\alpha}) - q_{1n}]}{p^1 \cdot q^1}$$

$$= Q^t_{\text{UV}\alpha} \varepsilon^t_{L\alpha}$$

where the period t error term $\varepsilon^t_{L\alpha}$ is defined for $t = 1, \ldots, T$ as[101]:

$$\varepsilon^t_{L\alpha} \equiv \sum_{n=1}^{N} \alpha_n \frac{\left[p^1_{\text{UV}\alpha} - \frac{p_{1n}}{\alpha_n} \right] \left[\frac{q_{tn}}{Q^t_{\text{UV}\alpha}} - q_{1n} \right]}{p^1 \cdot q^1}. \tag{11.108}$$

[101] This error term is homogeneous of degree 0 in the components of p^1, q^1 and q^t. Hence it is invariant to proportional changes in the components of these vectors. Definition (11.108) is only valid if all $\alpha_n > 0$. If this is not the case, redefine $\varepsilon^t_{L\alpha}$ as $\sum_{n=1}^{N} \left[\alpha_n p^1_{\text{UV}\alpha} - p_{1n} \right] [q_{tn} - q_{1n} Q^t_{\text{UV}\alpha}] \big/ p^1 \cdot q^1$. and with this change, the decomposition defined by the last line of (11.107) will continue to hold. It should be noted that $\varepsilon^t_{L\alpha}$ does not have an interpretation as a *covariance* between a vector of price differences and a vector of quantity differences.

Assuming that $\alpha_n > 0$ for $n = 1, \ldots, N$, the vector of period t *quality adjusted prices* p^t_α is defined as follows for $t = 1, \ldots, T$:

$$p^t_\alpha \equiv [p_{t1\alpha}, \ldots, p_{tN\alpha}] \equiv \left[\frac{p_{t1}}{\alpha_1}, \frac{p_{t2}}{\alpha_2}, \ldots, \frac{p_{tN}}{\alpha_N}\right]. \tag{11.109}$$

It can be seen that $p^1_{\text{UV}\alpha} - (p_{1n}/\alpha_n)$ is the difference between the period 1 unit value price level, $p^1_{\text{UV}\alpha}$, and the period 1 quality adjusted price for product n, p_{1n}/α_n. Define the period t *quality adjusted quantity share for product* n (using the vector α of quality adjustment factors) as follows for $t = 1, \ldots, T$ and $n = 1, \ldots, N$:

$$S_{tn\alpha} \equiv \frac{\alpha_n q_{tn}}{\alpha \cdot q^t}. \tag{11.110}$$

The vector of *period t quality adjusted real product shares* (using the vector α of quality adjustment factors) is defined as $S^t_\alpha \equiv [S_{t1\alpha}, S_{t2\alpha}, \ldots, S_{tN\alpha}]$ for $t = 1, \ldots, T$. It can be seen that these vectors are share vectors in that their components sum to 1; i.e., we have for $t = 1, \ldots, T$:

$$1_N \cdot S^t_\alpha = 1. \tag{11.111}$$

Using the above definitions, we can show that the period t quality adjusted unit value price level, $p^t_{\text{UV}\alpha}$ defined by (11.104) is equal to a share weighted average of the period t quality adjusted prices $p_{tn\alpha} = p_{tn}/\alpha_n$ defined by (11.109); i.e., for $t = 1, \ldots, T$, we have the following equations:

$$p^t_{\text{UV}\alpha} = p^t \cdot q^t / \alpha \cdot q^t \tag{11.112}$$

using (11.104)

$$= \sum_{n=1}^{N} \frac{(p_{tn}/\alpha_n)(\alpha_n q_{tn})}{\alpha \cdot q^t}$$

$$= \sum_{n=1}^{N} S_{tn\alpha} p_{tn\alpha}$$

using (11.109) and (11.110)

$$= S^t_\alpha \cdot p^t_\alpha.$$

Now we are in a position to determine the likely sign of $\varepsilon_{L\alpha}^{t}$ defined by (11.108). If the products are substitutable, it is likely that $\varepsilon_{L\alpha}^{t}$ is *negative*, since if p_{1n} is unusually low, then it is likely that the quality adjusted price for product n, p_{1n}/α_n, is below the weighted average of the quality adjusted prices for period 1 which is $p_{\text{UV}\alpha}^{1} = S_{\alpha}^{1} \cdot p_{\alpha}^{1}$ using (11.112) for $t = 1$. Thus we expect that $p_{\text{UV}\alpha}^{1} - (p_{1n}/\alpha_n) > 0$. If p_{1n} is unusually low, then we would expect that the corresponding q_{1n} is unusually high, and thus it is likely that q_{1n} is greater than $q_{tn}/Q_{\text{UV}\alpha}^{t}$ and so $q_{tn}/Q_{\text{UV}\alpha}^{t} - q_{1n} < 0$. Thus the sum of the N terms on the right hand side of (11.108) is likely to be negative. Our expectation[102] is that the error term $\varepsilon_{L\alpha}^{t} < 0$ and hence $Q_{\text{UV}\alpha}^{t} < Q_{L}^{t}$ for $t \geq 2$.

As usual, there are three special cases of (11.108) that will imply that $Q_{\text{UV}\alpha}^{t} = Q_{L}^{t}$: (i) $p_{\alpha}^{1} = \lambda_1 1_N$ so that all period 1 quality adjusted prices are equal[103]; (ii) $q^t = \lambda_t q^1$ for $t = 2, 3, \ldots, T$ so that quantities vary in strict proportion over time; (iii) the following sum of price differences times quantity differences equals 0; i.e., $\sum_{n=1}^{N} [\alpha_n p_{\text{UV}\alpha}^{1} - p_{1n}][(q_{tn}/Q_{\text{UV}\alpha}^{t}) - q_{1n}] = 0$.

If we divide both sides of equation t in Eq. (11.108) by $Q_{\text{UV}\alpha}^{t}$, we obtain the following system of identities for $t = 1, \ldots, T$:

$$\frac{Q_{L}^{t}}{Q_{\text{UV}\alpha}^{t}} = 1 - \varepsilon_{L\alpha}^{t} \qquad (11.113)$$

where we expect $\varepsilon_{L\alpha}^{t}$ to be a small negative number if the products are substitutes.[104]

The difference between the reciprocal of the quality adjusted unit value quantity index for period t, $[Q_{UV\alpha}^{t}]^{-1}$ and the reciprocal of the Paasche quantity index for period t, $[Q_{P}^{t}]^{-1}$, can be written as follows for $t = 1$,

[102] As in the previous section, this expectation is not held with great conviction if the period t quantities have a large variance.

[103] The condition $p_{\alpha}^{t} = \lambda_1 1_N$ is equivalent to $p^1 = \lambda_1 \alpha$. Thus if we choose α to be proportional to the period 1 price vector p^1, then $Q_{\text{UV}\alpha}^{t} = Q_{L}^{t}$ and $P_{\text{UV}\alpha}^{t} = P_{P}^{t}$, the fixed base Paasche price index. Thus, with this choice of α, the quality adjusted unit value index will usually have a downward bias relative to a superlative index. This result requires that p^1 be strictly positive.

[104] If $q_{1n} = 0$ and the period 1 quality adjusted reservation price p_{1n}/α_n is greater than the period 1 unit value price $P_{\text{UV}\alpha}^{1}$, then $\varepsilon_{L\alpha}^{t}$ defined by (11.108) could be a large negative number.

$\ldots, T$:

$$[Q^t_{\mathrm{UV}\alpha}]^{-1} - [Q^t_P]^{-1} = \left[\frac{\alpha \cdot q^1}{\alpha \cdot q^t}\right] - \left[\frac{p^t \cdot q^1}{p^t \cdot q^t}\right] \tag{11.114}$$

using (11.84) and (11.103)

$$= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n q_{1n}}{\alpha \cdot q^t}\right) - \left(\frac{p_{tn} q_{1n}}{p^t \cdot q^t}\right)\right]$$

$$= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n}{\alpha \cdot q^t}\right) - \left(\frac{p_{tn}}{p^t \cdot q^t}\right)\right] q_{1n}$$

$$= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n p^t \cdot q^t}{\alpha \cdot q^t}\right) - p_{tn}\right]\left[\frac{q_{1n}}{p^t \cdot q^t}\right]$$

$$= \sum_{n=1}^{N} [\alpha_n p^t_{\mathrm{UV}\alpha} - p_{tn}]\left[\frac{q_{1n}}{p^t \cdot q^t}\right]$$

using (11.104)

$$= \sum_{n=1}^{N} \frac{[\alpha_n p^t_{\mathrm{UV}\alpha} - p_{tn}]\left[q_{1n} - \left(\frac{q_{tn}}{Q^t_{\mathrm{UV}\alpha}}\right)\right]}{p^t \cdot q^t}$$

using (11.106)

$$= [Q^t_{\mathrm{UV}\alpha}]^{-1} \sum_{n=1}^{N} \alpha_n \frac{[p^t_{\mathrm{UV}\alpha} - (p_{tn}/\alpha_n)][(q_{1n} Q^t_{\mathrm{UV}\alpha} - q_{tn})]}{p^t \cdot q^t}$$

$$= [Q^t_{\mathrm{UV}\alpha}]^{-1} \varepsilon^t_{P\alpha}$$

where the period t error term $\varepsilon^t_{P\alpha}$ is defined for $t = 1, \ldots, T$ as[105]:

$$\varepsilon^t_{P\alpha} \equiv \sum_{n=1}^{N} \alpha_n \frac{[p^t_{\text{UV}\alpha} - (p_{tn}/\alpha_n)][(q_{1n} Q^t_{\text{UV}\alpha}) - q_{tn}]}{p^t \cdot q^t}. \tag{11.115}$$

If the products are substitutable, it is likely that $\varepsilon^t_{P\alpha}$ is *negative*, since if p_{tn} is unusually low, then it is likely that the period t quality adjusted price for product n, p_{tn}/α_n, is below the weighted average of the quality adjusted prices for period t which is $p^t_{\text{UV}\alpha} = S^t_\alpha p^t_\alpha$ using (11.112). Thus we expect that $p^t_{\text{UV}\alpha} - (p_{tn}/\alpha_n) > 0$. If p_{tn} is unusually low, then we would expect that the corresponding q_{tn} is unusually high, and thus it is likely that q_{tn} is greater than $q_{1n} Q^t_{\text{UV}\alpha}$ and so $q_{1n} Q^t_{\text{UV}\alpha} - q_{tn} < 0$. Thus the sum of the N terms on the right hand side of (11.115) is likely to be negative. Thus our expectation is that the error term $\varepsilon^t_{P\alpha} < 0$ and hence $[Q^t_{\text{UV}\alpha}]^{-1} < [Q^t_P]^{-1}$ for $t \geq 2$. Assuming that $\varepsilon^t_{L\alpha}$ is also negative, we have $Q^t_P < Q^t_{\text{UV}\alpha} < Q^t_L$ for $t = 2, \ldots, T$ as inequalities that are likely to hold.

As usual, there are three special cases of (11.114) that will imply that $Q^t_{\text{UV}\alpha} = Q^t_P$: (i) $p^t_\alpha = \lambda_t 1_N$ so that all period t quality adjusted prices are equal; (ii) $q^t = \lambda_t q^1$ for $t = 2, 3, \ldots, T$ so that quantities vary in strict proportion over time; (iii) the following sum of price differences times quantity differences equals zero: i.e., $\sum_{n=1}^{N} [\alpha_n p^t_{\text{UV}\alpha} - p_{tn}][(q_{1n} Q^t_{\text{UV}\alpha} - q_{tn})] = 0$.

If we divide both sides of equation t in Eq. (11.114) by $[Q^t_{\text{UV}\alpha}]^{-1}$, we obtain the following system of identities for $t = 1, \ldots, T$:

$$\frac{Q^t_P}{Q^t_{UV\alpha}} = [1 - \varepsilon^t_{P\alpha}]^{-1} \tag{11.116}$$

where we expect $\varepsilon^t_{P\alpha}$ to be a small negative number if the products are substitutes.

Equations (11.113) and (11.116) develop exact relationships for the quality adjusted unit value quantity index $Q^t_{\text{UV}\alpha}$ with the corresponding

[105] This error term is homogeneous of degree 0 in the components of p^t, q^1 and q^t. Hence, it is invariant to proportional changes in the components of these vectors. Definition (11.115) is only valid if all $\alpha_n > 0$. If this is not the case, then redefine $\varepsilon^t_{P\alpha}$ as $\sum_{n=1}^{N} [\alpha_n P^t_{\text{UV}\alpha} - p_{tn}][q_{1n} Q^t_{\text{UV}\alpha} - q_{tn}]/p^t \cdot q^t$ and with this change, the decomposition defined by the last line of (11.114) will continue to hold.

fixed base Laspeyres and Paasche quantity indexes, Q^t_L and Q^t_P. Taking the square root of the product of these two sets of equations leads to the following exact relationships between the fixed base Fisher quantity index, Q^t_F, and its quality adjusted unit value counterpart period t quantity index, $Q^t_{\mathrm{UV}\alpha}$, for $t = 1, \ldots, T$:

$$Q^t_F = Q^t_{\mathrm{UV}\alpha}\left[\frac{(1-\varepsilon^t_{L\alpha})}{(1-\varepsilon^t_{P\alpha})}\right]^{1/2} \tag{11.117}$$

where $\varepsilon^t_{L\alpha}$ and $\varepsilon^t_{P\alpha}$ are defined by (11.108) and (11.115). If there are no strong (divergent) trends in prices and quantities, then it is likely that $\varepsilon^t_{L\alpha}$ is approximately equal to $\varepsilon^t_{P\alpha}$ and hence under these conditions, it is likely that $Q^t_{\mathrm{UV}\alpha} \approx Q^t_F$; i.e., the quality adjusted unit value quantity index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, if there are divergent trends in prices and quantities (in opposite directions), then it is likely that $\varepsilon^t_{P\alpha}$ will be more negative than $\varepsilon^t_{L\alpha}$ and hence it is likely that $Q^t_F < Q^t_{\mathrm{UV}\alpha}$ for $t = 2, \ldots, T$; i.e., *with divergent trends in prices and quantities, the quality adjusted unit value quantity index is likely to have an upward bias* relative to its Fisher quantity index counterparts.[106]

Using Eq. (11.105), we have the following counterparts to Eq. (11.94) for $t = 1, \ldots, T$:

$$\frac{p^t \cdot q^t}{p^1 \cdot q^1} = P^t_{\mathrm{UV}\alpha} Q^t_{\mathrm{UV}\alpha} = P^t_P Q^t_L = P^t_L Q^t_P. \tag{11.118}$$

Equations (11.113), (11.116) and (11.118) imply the following identities for $t = 1, \ldots, T$:

$$\frac{P^t_{\mathrm{UV}\alpha}}{P^t_P} = \frac{Q^t_L}{Q^t_{\mathrm{UV}\alpha}} = 1 - \varepsilon^t_{L\alpha}; \tag{11.119}$$

$$\frac{P^t_{\mathrm{UV}\alpha}}{P^t_L} = \frac{Q^t_P}{Q^t_{\mathrm{UV}\alpha}} = \left[1 - \varepsilon^t_{P\alpha}\right]^{-1}; \tag{11.120}$$

[106] As was the case in the previous section, if there are missing products in period 1, the expected inequality $Q^t_F < Q^t_{\mathrm{UV}\alpha}$ may be reversed, because $\varepsilon^t_{L\alpha}$ defined by (11.108) may become significantly negative if some q_{1n} equal 0 while their corresponding reservation prices p_{1n} are positive.

We expect that ε_L^t and $\varepsilon_{P\alpha}^t$ will be predominantly negative if the products are highly substitutable and thus in this case, the quality adjusted unit value indexes $P_{\text{UV}\alpha}^t$ should satisfy the inequalities $P_P^t < P_{\text{UV}\alpha}^t < P_L^t$ for $t = 2, 3, \ldots, T$..

Taking the square root of the product of Eqs. (11.119) and (11.120) leads to the following exact relationships between the fixed base Fisher price index, P_F^t, and its quality adjusted unit value counterpart period t index, $P_{\text{UV}\alpha}^t$, for $t = 1, \ldots, T$:

$$P_{\text{UV}\alpha}^t = P_F^t \left\{ \frac{\left(1 - \varepsilon_{L\alpha}^t\right)}{\left(1 - \varepsilon_{P\alpha}^t\right)} \right\}^{1/2} \tag{11.121}$$

where $\varepsilon_{L\alpha}^t$ and $\varepsilon_{P\alpha}^t$ are defined by (11.108) and (11.115). If there are no strong (divergent) trends in prices and quantities, then it is likely that $\varepsilon_{L\alpha}^t$ is approximately equal to $\varepsilon_{P\alpha}^t$ and hence under these conditions, it is likely that $P_{\text{UV}\alpha}^t \approx P_F^t$; i.e., the quality adjusted unit value price index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, if there are divergent trends in prices and quantities, then we expect $\varepsilon_{P\alpha}^t$ to be more negative than $\varepsilon_{L\alpha}^t$ and hence there is an expectation that $P_{\text{UV}\alpha}^t < P_F^t$ for $t = 2, \ldots, T$; i.e., we expect that normally $P_{\text{UV}\alpha}^t$ will have a *downward bias* relative to P_F^t.[107] However, if there are missing products in period 1, then the bias of $P_{\text{UV}\alpha}^t$ relative to P_F^t is uncertain.

Relationships Between Lowe and Fisher Indexes

We now consider how a Lowe (1823) price index is related to a fixed base Fisher price index. The framework that we consider is similar to the framework developed in section "Relationships Between Superlative Fixed Base Indexes and Geometric Indexes That Use Average Annual Shares as Weights" for the annual share weighted Jevons index, $P_{J\alpha}^t$. In the present section, instead of using the average sales shares for the first year in the

[107] Recall that the weighted unit value quantity level, $q_{\text{UV}\alpha}^t$ is defined as the linear function of the period t quantity data, $\alpha \cdot q^t$. If $T \geq 3$ and the price and quantity data are consistent with purchasers maximizing a utility function that generates data that is exact for the Fisher price index Q_F^t, then $Q_{\text{UV}\alpha}^t$ will tend to be greater than Q_F^t (and hence $P_{\text{UV}\alpha}^t$ will tend to be less than P_F^t) for $t \geq 2$. See Marris (1984, 52), Diewert (1999b, 49) and Diewert and Fox (2020) on this point.

sample as weights for a weighted Jevons index, we use annual average quantities sold (or purchased) in the first year as a vector of quantity weights for subsequent periods. Define the *annual average quantity vector* $q^* \equiv [q_1^*, \ldots, q_N^*]$ for the first T^* periods in the sample that make up a year, q^*, as follows[108]:

$$q^* \equiv \left(\frac{1}{T^*}\right) \sum_{t=1}^{T^*} q^t. \tag{11.122}$$

As was the case in section "Relationships Between Superlative Fixed Base Indexes and Geometric Indexes That Use Average Annual Shares as Weights", the reference year for the weights precedes the reference month for the product prices. Define the *period t* Lowe (1823) *price level* and *price index*, p_{Lo}^t and P_{Lo}^t by (11.123) and (11.124), respectively, for $t = T^* + 1, T^* + 2, \ldots, T$:

$$p_{\text{Lo}}^t \equiv p^t \cdot \alpha; \tag{11.123}$$

$$P_{\text{Lo}}^t \equiv \frac{p_{\text{Lo}}^t}{p_{\text{Lo}}^{T^*+1}} = \frac{p^t \cdot \alpha}{p^{T^*+1} \cdot \alpha} \tag{11.124}$$

where the constant price weights vector α is the annual average weights vector q^* defined by (11.122); i.e., we have:

$$\alpha \equiv q^*. \tag{11.125}$$

The *period t Lowe quantity level*, q_{Lo}^t, and the corresponding *period t Lowe quantity index*, Q_{Lo}^t, are defined as follows for $t = T^* + 1, T^* + 2,$

[108] If product n was not available in the first year of the sample, then the nth component of q^*, q_n^*, will equal 0 and hence the nth component of the weight vector α defined by (11.125) will also equal 0. If product n was also not available in periods $t \geq T^* + 1$, then looking at definitions (11.123) and (11.124), it can be seen that P_{Lo}^t will not depend on the reservation prices p_{nt} for these subsequent periods where product n is not available. Thus, under these circumstances, the Lowe index cannot be consistent with the (Hicksian) economic approach to index number theory since Konüs (1924) true cost of living price indexes will depend on the reservation prices. However, if the products in the elementary aggregate are indeed highly substitutable, then the assumption of a linear utility function will provide an adequate approximation to the "truth" and the estimation of reservation prices becomes unimportant.

$\ldots, T$[109]:

$$q_{\text{Lo}}^t \equiv \frac{p^t \cdot q^t}{p_{\text{Lo}}^t} = \frac{p^t \cdot q^t}{p^t \cdot \alpha} = \sum_{n=1}^{N} \left(\frac{p_{tn}\alpha_n}{p^t \cdot \alpha}\right)\left(\frac{q_{tn}}{\alpha_n}\right); \tag{11.126}$$

$$Q_{\text{Lo}}^t \equiv \frac{q_{\text{Lo}}^t}{q_{\text{Lo}}^{T^*+1}} = \left[\frac{p^t \cdot q^t \Big/ p^{T^*+1} \cdot q^{T^*+1}}{P_{\text{Lo}}^t}\right]. \tag{11.127}$$

It can be seen that the Lowe price index defined by (11.124) is equal to a *weighted Dutot price index*; see definition (11.14) above. It is also structurally identical to the quality adjusted unit value quantity index $Q_{\text{UV}\alpha}^t$ defined in the previous section, except the role of prices and quantities has been reversed. Thus the identity (11.107) in the previous section will be valid if we replace $Q_{UV\alpha}^t$ by P_{Lo}^t, replace Q_L^t by P_L^t and interchange prices and quantities on the right hand side of (11.107).[110] The resulting identities are the following ones for $t = T^* + 1, T^* + 2, \ldots, T$[111]:

$$\begin{aligned} P_{\text{Lo}}^t - P_L^t &= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n p_{tn}}{\alpha \cdot p^{T^*+1}}\right) - \left(\frac{p_{tn} q_{T^*+1,n}}{p^{T^*+1} \cdot q^{T^*+1}}\right)\right] \\ &= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n}{\alpha \cdot p^{T^*+1}}\right) - \left(\frac{q_{T^*+1,n}}{p^{T^*+1} \cdot q^{T^*+1}}\right)\right] p_{tn} \\ &= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n p^{T^*+1} \cdot q^{T^*+1}}{\alpha \cdot P^{T^*+1}}\right) - q_{T^*+1,n}\right]\left[\frac{p_{tn}}{p^{T^*+1} \cdot q^{T^*+1}}\right] \\ &= \sum_{n=1}^{N} \left[\alpha_n q_{\text{Lo}}^{T^*+1} - q_{T^*+1,n}\right]\left[\frac{p_{tn}}{p^{T^*+1} \cdot q^{T^*+1}}\right] \end{aligned} \tag{11.128}$$

[109] This last inequality is only valid if all $\alpha_n > 0$. It can be seen that the Lowe quantity level for period t, q_{Lo}^t, is a share weighted sum of the period t quality adjusted quantities, q_{tn}/α_n.

[110] We also replace period 1 by period $T^* + 1$.

[111] This step follows using the following counterpart to (11.106): $\sum_{n=1}^{N} \left[\alpha_n q_{\text{Lo}}^{T^*+1} - q_{T^*+1,n}\right] p_{T^*+1,n} = 0$.

using (11.126) for $t = T^* + 1$

$$
\begin{aligned}
&= \sum_{n=1}^{N} \frac{\left[\alpha_n q_{\text{Lo}}^{T^*+1} - q_{T^*+1,n}\right]\left[p_{tn} - p_{T^*+1,n} P_{\text{Lo}}^t\right]}{p^{T^*+1} \cdot q^{T^*+1}} \\
&= P_{\text{Lo}}^t \sum_{n=1}^{N} \frac{\left[\alpha_n q_{\text{Lo}}^{T^*+1} - q_{T^*+1,n}\right]\left[\left(\frac{p_{tn}}{P_{\text{Lo}}^t}\right) - p_{T^*+1,n}\right]}{p^{T^*+1} \cdot q^{T^*+1}} \\
&= P_{\text{Lo}}^t \sum_{n=1}^{N} \alpha_n \frac{\left[q_{\text{Lo}}^{T^*+1} - \left(q_{T^*+1,n}/\alpha_n\right)\right]\left[\left(p_{tn}/P_{\text{Lo}}^t\right) - p_{T^*+1,n}\right]}{p^{T^*+1} \cdot q^{T^*+1}} \\
&= P_{\text{Lo}}^t \varepsilon_{L\alpha}^t
\end{aligned}
$$

where the period t error term $\varepsilon_{L\alpha}^t$ is now defined for $t = T^* + 1, \ldots, T$ as follows[112]:

$$
\varepsilon_{L\alpha}^t \equiv \sum_{n=1}^{N} \alpha_n \frac{\left[q_{\text{Lo}}^{T^*+1} - \left(q_{T^*+1,n}/\alpha_n\right)\right]\left[\left(p_{tn}/P_{\text{Lo}}^t\right) - p_{T^*+1,n}\right]}{p^{T^*+1} \cdot q^{T^*+1}}. \qquad (11.129)
$$

If the products are substitutable, it is likely that $\varepsilon_{L\alpha}^t$ is *negative*, since if $p_{T^*+1,n}$ is unusually low, then it is likely that $\left(p_{tn}/P_{\text{Lo}}^t\right) - p_{T^*+1,n} > 0$ and that $q_{T^*+1,n}/\alpha_n$ is unusually large and hence is greater than $q_{\text{Lo}}^{T^*+1}$, which is a weighted average of the period $T^* + 1$ quantity ratios, $q_{T^*+1,1}/\alpha_1$, $q_{T^*+1,2}/\alpha_2, \ldots, q_{T^*+1,N}/\alpha_N$ using definition (11.126) for $t = T^* + 1$. Thus the sum of the N terms on the right hand side of (11.129) is likely to be negative. Thus, our expectation[113] is that the error term $\varepsilon_{L\alpha}^t < 0$ and hence $P_{\text{Lo}}^t < P_L^t$ for $t > T^* + 1$.

The α_n can be interpreted as *inverse quality indicators* of the utility provided by one unit of the nth product. Suppose purchasers

[112] Note that this error term is homogeneous of degree 0 in the components of p^{T^*+1}, q^{T^*+1} and p^t. Hence, it is invariant to proportional changes in the components of these vectors. Definition (11.129) is only valid if all $\alpha_n > 0$. If this is not the case, redefine $\varepsilon_{L\alpha}^t$ as $\sum_{n=1}^{N}\left[\alpha_n q_{Lo}^{T^*+1} - q_{T^*+1,n}\right]\left[\left(p_{tn}/P_{Lo}^t\right) - p_{T^*+1,n}\right]/\,p^{T^*+1} \cdot q^{T^*+1}$ and with this change, the decomposition defined by the last line of (11.128) will continue to hold.

[113] This expectation is not held with great conviction if the period t prices have a large variance.

of the N commodities have Leontief preferences with the utility function $f(q_1, q_2, \ldots, q_N) \equiv \min_n\{q_n/\alpha_n : n = 1, 2, \ldots N\}$. Then the dual unit cost function that corresponds to this functional form is $c(p_1, p_2, \ldots, p_N) \equiv \sum_{n=1}^{N} p_n\alpha_n = p\cdot\alpha$. If we evaluate the unit cost function at the prices of period t, p^t, we obtain the Lowe price level for period t defined by (11.123); i.e., $p^t_{\text{Lo}} \equiv p^t \cdot \alpha$. Thus the bigger α_n is, the more units of q_n it will take for purchasers of the N commodities to attain one unit of utility. Thus the α_n can be interpreted as inverse indicators of the relative utility of each product.

As usual, there are three special cases of (11.128) that will imply that $P^t_{\text{Lo}} = P^t_L$: (i) $q^{T^*+1} = \lambda q^*$ for some $\lambda > 0$ so that the period $T^* + 1$ quantity vector q^{T^*+1} is proportional to the annual average quantity vector q^* for the base year; (ii) $p^t = \lambda_t p^{T^*+1}$ for some $\lambda_t > 0$ for $t = T^* + 1, \ldots, T$ so that prices vary in strict proportion over time; (iii) the sum of terms $\sum_{n=1}^{N}\left[\alpha_n q^{T^*+1}_{\text{Lo}} - q_{T^*+1,n}\right]\left[\left(p_{tn}/P^t_{\text{Lo}}\right) - p_{T^*+1,n}\right] = 0$.

If we divide both sides of equation t in Eq. (11.128) by P^t_{Lo}, we obtain the following system of identities for $t = T^* + 1, \ldots, T$:

$$\frac{P^t_L}{P^t_{\text{Lo}}} = 1 - \varepsilon^t_{L\alpha} \tag{11.130}$$

where we expect $\varepsilon^t_{L\alpha}$ to be a small negative number.

We turn now to developing a relationship between the Lowe and Paasche price indexes. The difference between reciprocal of the Lowe price index for period t, $\left[P^t_{\text{Lo}}\right]^{-1}$ and the reciprocal of the Paasche price index for period t, $\left[P^t_P\right]^{-1}$, can be written as follows for $t = T^* + 1, \ldots, T$:

$$\begin{aligned}
\left[P^t_{\text{Lo}}\right]^{-1} - \left[P^t_P\right]^{-1} &= \left[\frac{\alpha \cdot p^{T^*+1}}{\alpha \cdot p^t}\right] - \left[\frac{q^t \cdot p^{T^*+1}}{q^t \cdot p^t}\right] \\
&= \sum_{n=1}^{N}\left[\left(\frac{\alpha_n p_{T^*+1,n}}{\alpha \cdot p^t}\right) - \left(\frac{q_{tn} p_{T^*+1,n}}{p^t \cdot q^t}\right)\right] \\
&= \sum_{n=1}^{N}\left[\left(\frac{\alpha_n}{\alpha \cdot p^t}\right) - \left(\frac{q_{tn}}{p^t \cdot q^t}\right)\right] p_{T^*+1,n}
\end{aligned}$$

$$= \sum_{n=1}^{N} \left[\left(\frac{\alpha_n p^t \cdot q^t}{\alpha \cdot p^t} \right) - q_{tn} \right] \left[\frac{p_{T^*+1,n}}{p^t \cdot q^t} \right]$$

$$= \sum_{n=1}^{N} [\alpha_n q^t_{\mathrm{Lo}} - q_{tn}] \left[\frac{p_{T^*+1,n}}{p^t \cdot q^t} \right] \quad (11.131)$$

using (11.126)[114]

$$= \sum_{n=1}^{N} \frac{[\alpha_n q^t_{Lo} - q_{tn}] \left[p_{T^*+1,n} - \frac{p_{tn}}{P^t_{Lo}} \right]}{p^t \cdot q^t}$$

$$= [P^t_{Lo}]^{-1} \sum_{n=1}^{N} \frac{[\alpha_n q^t_{Lo} - q_{tn}][p_{T^*+1,n} P^t_{Lo} - p_{tn}]}{p^t \cdot q^t}$$

$$= [P^t_{Lo}]^{-1} \sum_{n=1}^{N} \alpha_n \frac{[q^t_{Lo} - (q_{tn}/\alpha_n)][p_{T^*+1,n} P^t_{Lo} - p_{tn}]}{p^t \cdot q^t} \text{ if all } \alpha_n > 0$$

$$= [P^t_{Lo}]^{-1} \varepsilon^t_{P\alpha}$$

where the period t error term $\varepsilon^t_{P\alpha}$ is defined for $t = T^* + 1, \ldots, T$ as[115]:

$$\varepsilon^t_{P\alpha} \equiv \sum_{n=1}^{N} \alpha_n \frac{[q^t_{\mathrm{Lo}} - (q_{tn}/\alpha_n)][p_{T^*+1,n} P^t_{\mathrm{Lo}} - p_{tn}]}{p^t \cdot q^t}. \quad (11.132)$$

If the products are substitutable, it is likely that $\varepsilon^t_{P\alpha}$ is *negative*, since if p_{tn} is unusually low, then it is likely that it will be less than the inflation adjusted n-th component of the period $T^* + 1$ price, $p_{T^*+1,n} P^t_{\mathrm{Lo}}$. If p_{tn} is unusually low, then it is also likely that the period t quality adjusted quantity for product n, q_{tn}/α_n, is above the weighted average of the quality adjusted quantities for period t which is q^t_{Lo}. Thus the sum of

[114] This step follows using the following counterpart to (11.106): $\sum_{n=1}^{N} [\alpha_n q^t_{\mathrm{Lo}} - q_{tn}] p_{tn} = 0$.

[115] This error term is homogeneous of degree 0 in the components of q^t, p^{T^*+1} and p^t. Hence, it is invariant to proportional changes in the components of these vectors. Definition (11.132) is only valid if all $\alpha_n > 0$. If this is not the case, redefine $\varepsilon^t_{P\alpha}$ as $\sum_{n=1}^{N} [\alpha_n q^t_{\mathrm{Lo}} - q_{tn}][p_{T^*+1,n} P^t_{\mathrm{Lo}} - p_{tn}]/ p^t \cdot q^t$ and with this change, the decomposition defined by the last line of (11.131) will continue to hold.

the N terms on the right hand side of (11.132) is likely to be negative. Thus our expectation is that the error term $\varepsilon^t_{P\alpha} < 0$ and hence $\left[P^t_{\mathrm{Lo}}\right]^{-1} < \left[P^t_P\right]^{-1}$ for $t = T^*+2, \ldots, T$. Assuming that $\varepsilon^t_{L\alpha}$ is also negative, we have $P^t_P < P^t_{\mathrm{Lo}} < P^t_L$ for $t = T^*+2, T^*+3, \ldots, T$ as inequalities that are likely to hold.

As usual, there are three special cases of (11.131) that will imply that $P^t_{\mathrm{Lo}} = P^t_P$: (i) $q^t = \lambda q^*$ for some $\lambda > 0$ so that the period t quantity vector q^t is proportional to the annual average quantity vector q^* for the reference year prior to the reference month; (ii) $p^t = \lambda_t p^{T^*+1}$ for $t = T^*+2, T^*+3, \ldots, T$ so that prices vary in strict proportion over time; (iii) the sum of terms $\sum_{n=1}^{N}\left[\alpha_n q^t_{\mathrm{Lo}} - q_{tn}\right]\left[p_{T^*+1,n} P^t_{\mathrm{Lo}} - p_{tn}\right] = 0$.

If we divide both sides of equation t in Eq. (11.131) by $\left[P^t_{\mathrm{Lo}}\right]^{-1}$, we obtain the following system of identities for $t = T^*+1, \ldots, T$:

$$\frac{P^t_P}{P^t_{\mathrm{Lo}}} = \left[1 - \varepsilon^t_{P\alpha}\right]^{-1} \tag{11.133}$$

where we expect $\varepsilon^t_{P\alpha}$ to be a negative number.

Equations (11.130) and (11.133) develop exact relationships for the Lowe price index P^t_{Lo} with the corresponding fixed base Laspeyres and Paasche price indexes, P^t_L and P^t_P. Taking the square root of the product of these two sets of equations leads to the following exact relationships between the fixed base Fisher price index, P^t_F, and the corresponding Lowe period t price index, P^t_{Lo}, for $t = T^*+1, \ldots, T$:

$$P^t_F = P^t_{\mathrm{Lo}}\left\{\frac{\left(1-\varepsilon^t_{L\alpha}\right)}{\left(1-\varepsilon^t_{P\alpha}\right)}\right\}^{1/2} \tag{11.134}$$

where $\varepsilon^t_{L\alpha}$ and $\varepsilon^t_{P\alpha}$ are defined by (11.129) and (11.132). If there are no strong (divergent) trends in prices and quantities, then it is likely that $\varepsilon^t_{L\alpha}$ is approximately equal to $\varepsilon^t_{P\alpha}$ and hence under these conditions, it is likely that $P^t_{\mathrm{Lo}} \approx P^t_F$; i.e., the Lowe price index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, if there are divergent trends in prices and quantities (in diverging directions), then it is likely that $\varepsilon^t_{P\alpha}$ will be more negative than $\varepsilon^t_{L\alpha}$ and hence it is likely that $P^t_F < P^t_{\mathrm{Lo}}$ for $t = T^*+2, \ldots, T$; i.e., *with divergent trends in prices and quantities, the Lowe price index is likely to have an upward bias* relative to its Fisher Price index counterpart. This is an intuitively plausible result since the Lowe index is a fixed basket type index and hence will be subject to some upward substitution bias relative to the Fisher index which is able to control for substitution bias.

In the following section, we show that the Geary Khamis multilateral indexes can be regarded as quality adjusted unit value price indexes and hence the analysis in section "Quality Adjusted Unit Value Price and Quantity Indexes" on quality adjusted unit value price indexes can be applied to GK multilateral indexes.

Geary Khamis Multilateral Indexes

The GK multilateral method was introduced by Geary (1958) in the context of making international comparisons of prices. Khamis (1970) showed that the equations that define the method have a positive solution under certain conditions. A modification of this method has been adapted to the time series context and is being used to construct some components of the Dutch CPI; see Chessa (2016). The GK index was the multilateral index chosen by the Dutch to avoid the chain drift problem for the segments of their CPI that use scanner data.

The GK system of equations for T time periods involves T *price levels* $p_{\mathrm{GK}}^1, \ldots, p_{\mathrm{GK}}^T$ and N *quality adjustment factors* $\alpha_1, \ldots, \alpha_N$.[116] Let p^t and q^t denote the N dimensional price and quantity vectors for period t (with components p_{tn} and q_{tn} as usual). Define the total consumption (or sales) vector q over the entire window of observations as the following simple sum of the period by period consumption vectors:

$$q \equiv \sum_{t=1}^{T} q^t \tag{11.135}$$

where $q \equiv [q_1, q_2, \ldots, q_N]$. The equations which determine the *GK price levels* $p_{\mathrm{GK}}^1 \ldots, p_{\mathrm{GK}}^T$ and *quality adjustment factors* $\alpha_1, \ldots, \alpha_N$ (up to a scalar multiple) are the following ones:

$$\alpha_n = \sum_{t=1}^{T} \left[\frac{q_{tn}}{q_n}\right]\left[\frac{p_{tn}}{p_{\mathrm{GK}}^t}\right]; \quad n = 1, \ldots, N; \tag{11.136}$$

[116] In the international context, the α_n are interpreted as international commodity reference prices.

$$p_{GK}^t = \frac{p^t \cdot q^t}{\alpha \cdot q^t} = \sum_{n=1}^{N} \left[\frac{\alpha_n q_{tn}}{\alpha \cdot q^t}\right]\left[\frac{p_{tn}}{\alpha_n}\right]; \quad t = 1, \ldots, T \tag{11.137}$$

where $\alpha \equiv [\alpha_1, \ldots, \alpha_N]$ is the vector of GK quality adjustment factors. The sample share of period t's purchases of commodity n in total sales of commodity n over all T periods can be defined as $S_{tn} \equiv q_{tn}/q_n$ for $n = 1, \ldots, N$ and $t = 1, \ldots, T$. Thus $\alpha_n \equiv \sum_{t=1}^{T} S_{tn}[p_{tn}/p_{GK}^t]$ is a (real) share weighted average of the period t inflation adjusted prices p_{tn}/p_{GK}^t for product n over all T periods. The period t quality adjusted sum of quantities sold is defined as the period t GK quantity level, $q_{GK}^t \equiv \alpha \cdot q^t = \sum_{n=1}^{N} \alpha_n q_{tn}$.[117] This period t quantity level is divided into the value of period t sales, $p^t \cdot q^t = \sum_{n=1}^{N} p_{tn} q_{tn}$, in order to obtain the period t GK price level, p_{GK}^t. Thus the GK price level for period t can be interpreted as a *quality adjusted unit value index* where the α_n act as the quality adjustment factors.

Note that the GK price level, p_{GK}^t defined by (11.137) *does not depend on the estimated reservation prices*; i.e., the definition of p_{GK}^t zeros out any reservation prices that are applied to missing products and thus $P_{GK}^t \equiv p_{GK}^t/p_{GK}^1$ also does not depend on reservation prices.[118] A related property of the GK price levels is the following one: if a product n^* is only available in a single period t^*, then the GK price levels p_{GK}^t do not depend on p_{n*t*} or q_{n*t*}.[119]

It can be seen that if a solution to Eqs. (11.136) and (11.137) exists, then if all of the period price levels p_{GK}^t are multiplied by a positive scalar

[117] Khamis (1972, 101) also derived this equation in the time series context.

[118] In Eqs. (11.136) and (11.137), each price p_{tn} always appears with the multiplicative factor q_{tn}. Thus if p_{tn} is an imputed price, it will always be multiplied by $q_{tn} = 0$ and thus any imputed price will have no impact on the α_n and p_{GK}^t. Thus this method fails Test 9 in section "The Axiomatic Approach to Multilateral Price Levels".

[119] Let product n^* be available only in period t^*. Using (11.136) for $n = n^*$, we have: (i) $\alpha_{n*} = p_{t*n*}/p_{GK}^{t*}$. Eq. (11.137) can be rewritten as follows: (ii) $p_{GK}^t \alpha \cdot q^t = p^t \cdot q^t$; $t = 1, \ldots, T$. Note that for $t \neq t^*$, these equations do not depend directly on α_{n*}, p_{t*n*} or q_{t*n*}. For period $t = t^*$, equation t^* in (11.137) can be written as: (iii) $p_{GK}^{t*}\left(\sum_{n \neq n*} \alpha_n q_{t*n} + \alpha_{n*} q_{t*n*}\right) = \left(\sum_{n \neq n*} p_{t*n} q_{t*n} + p_{t*n*} q_{t*n*}\right)$. Substitute (i) into (iii) and after some simplification, we find that $p_{GK}^{t*} = \sum_{n \neq n*} p_{t*n} q_{t*n} \Big/ \sum_{n \neq n*} \alpha_n q_{t*n}$. This proof is due to Claude Lamboray. Thus this method fails Test 8 in section "The Axiomatic Approach to Multilateral Price Levels".

λ say and all of the quality adjustment factors α_n are divided by the same λ, then another solution to (11.136) and (11.137) is obtained. Hence, the α_n and p^t_{GK} are only determined up to a scalar multiple and an additional normalization is required such as $p^1_{GK} = 1$ or $\alpha_1 = 1$ is required to determine a unique solution to the system of equations defined by (11.136) and (11.137).[120] It can also be shown that only $N + T - 1$ of the $N + T$ equations in (11.136) and (11.137) are independent.

Using the normalization $p^1_{GK} = 1$, it is straightforward to show that the GK price levels, p^t_{GK}, are invariant to changes in the units of measurement. Suppose we have a solution p^t_{GK} and α_n for $t = 1, \ldots, T$ and $n = 1, \ldots, N$ with $p^1_{GK} \equiv 1$. Let $\lambda_n > 0$ for $n = 1, \ldots, N$. Use these λ_n to measure prices and quantities in new units of measurement; i.e., define $p^*_{tn} \equiv \lambda_n p_{tn}$ and $q^*_{tn} \equiv (\lambda_n)^{-1} q_{tn}$ for $t = 1, \ldots, T$ and $n = 1, \ldots, N$. Now substitute these transformed prices and quantities into Eqs. (11.135)–(11.137). It is straightforward to show that the initial solution GK price levels, p^t_{GK}, along with new $\alpha^*_n \equiv \lambda_n \alpha_n$ also satisfy the new GK Eqs. (11.135)–(11.137).

A traditional method for obtaining a solution to (11.136) and (11.137) is to iterate between these equations. Thus set $\alpha = 1_N$, a vector of ones, and use Eq. (11.137) to obtain an initial sequence for the p^t_{GK}. Substitute these p^t_{GK} estimates into Eq. (11.136) and obtain α_n estimates. Substitute these α_n estimates into Eq. (11.137) and obtain a new sequence of p^t_{GK} estimates. Continue iterating between the two systems until convergence is achieved.

An alternative method is more efficient. Following Diewert (1999b, 26),[121] substitute Eq. (11.137) into Eq. (11.136) and after some simplification, obtain the following system of equations that will determine the components of the α vector:

$$[I_N - C]\alpha = 0_N \tag{11.138}$$

where I_N is the N by N identity matrix, 0_N is a vector of zeros of dimension N and the C matrix is defined as follows:

$$C \equiv \hat{q}^{-1} \sum_{t=1}^{T} s^t q^{tT}. \tag{11.139}$$

[120] See Diewert and Fox (2020) for various solution methods.

[121] See also Diewert and Fox (2020) for additional discussion on this solution method.

where $\hat{q}$ is an N by N diagonal matrix with the elements of the total window purchase vector q running down the main diagonal and $\hat{q}^{-1}$ denotes the inverse of this matrix, s^t is the period t expenditure share column vector, q^t is the column vector of quantities purchased during period t and q_{tn} is the nth element of the sample total q defined by (11.135).

The matrix $I_N - C$ is singular which implies that the N equations in (11.138) are not all independent. In particular, if the first $N-1$ equations in (11.138) are satisfied, then the last equation in (11.138) will also be satisfied. It can also be seen that the N equations in (11.138) are homogeneous of degree one in the components of the vector α. Thus to obtain a unique b solution to (11.138), set α_N equal to 1, drop the last equation in (11.138) and solve the remaining $N-1$ equations for $\alpha_1, \alpha_2, \ldots, \alpha_{N-1}$. Once the α_n are known, Eq. (11.137) can be used to determine the GK price levels, $p^t_{GK} = p^t \cdot q^t / \alpha \cdot q^t$ for $t = 1, \ldots, T$..

Using Eq. (11.137), it can be seen that the *GK price index for period t* (relative to period 1) is equal to $P^t_{GK} \equiv p^t_{GK}/p^1_{GK} = [p^t \cdot q^t/\alpha \cdot q^t]/[p^1 \cdot q^1/\alpha \cdot q^1]$ for $t = 1, \ldots, T$ and thus these indexes are *quality adjusted unit value price indexes* with a particular choice for the vector of quality adjustment factors α. Thus these indexes lead to corresponding *additive quantity levels* q^t_{GK} that correspond to the linear utility function, $f(q) \equiv \alpha \cdot q$.[122] As we saw in section "Quality Adjusted Unit Value Price and Quantity Indexes", this type of index can approximate the corresponding fixed base Fisher price index provided that there are no systematic divergent trends in prices and quantities. However, if there are diverging trends in prices and quantities (in opposite directions), then we expect the GK price indexes to be subject to some *substitution bias* with the expectation that the GK price index for period $t \geq 2$ to be somewhat *below* the corresponding Fisher fixed base price index. Thus we expect GK and quality adjusted unit value price indexes to *normally*

[122] Using the economic approach to index number theory, it can be seen that the GK price indexes will be exactly the correct price indexes to use if purchasers maximize utility using a common linear utility function. Diewert (1999b, 27) and Diewert and Fox (2020) show that the GK price indexes will also be exactly correct if purchasers maximize a Leontief no substitution utility function. These extreme cases are empirically unlikely. As was noted earlier in section "Quality Adjusted Unit Value Price and Quantity Indexes", Leontief preferences are not consistent with new and disappearing products.

have a downward bias relative to their Fisher and Törnqvist counterparts, provided that there are no missing products, the products are highly substitutable and there are divergent trends in prices and quantities. However, if there are missing products in period 1, then it is quite possible for the GK price indexes to have an upward bias relative to their Fisher fixed base counterparts, which, in principle, use reservation prices for the missing products.[123]

In the following five sections, we will study in some detail another popular method for making price level comparisons over multiple periods: the Weighted Time Product Dummy Multilateral Indexes. The general case with missing observations will be studied in section "Weighted Time Product Dummy Regressions: The General Case". It proves to be useful to consider simpler special cases of the method in sections "Time Product Dummy Regressions: The Case of No Missing Observations"–"Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations".

Time Product Dummy Regressions: The Case of no Missing Observations

In this section, it is assumed that price and quantity data for N products are available for T periods. As usual, let $p^t \equiv [p_{t1}, \ldots, p_{tN}]$ and $q^t = [q_{t1}, \ldots, q_{tN}]$ denote the price and quantity vectors for time periods $t = 1, \ldots, T$. In this section, it is assumed that there are no missing prices or quantities so that all NT prices and quantities are positive. We assume initially that purchasers of the N products maximize the following linear utility function $f(q)$ defined as follows:

$$f(q) = f(q_1, q_2, \ldots, q_N) \equiv \sum_{n=1}^{N} \alpha_n q_n = \alpha \cdot q \qquad (11.140)$$

[123] New products appear with some degree of regularity and so it is likely that there will be missing products in period 1 and this may reverse the "normal" inequality, $P_{\mathrm{GK}}^t < P_F^t$, as was the case for Diewert's (2018) scanner data set. This data set is used in the Appendix to this chapter. The GK index, like all indexes based on quality adjusted unit values, zeros out the effects of reservation prices for the missing products, whereas Fisher indexes can include the effects of reservation prices.

where the α_n are positive parameters, which can be interpreted as quality adjustment parameters. Under the assumption of maximizing behavior on the part of purchasers of the N commodities, Wold's Identity[124] applied to a linearly homogeneous utility function tells us that the purchasers' system of *inverse demand functions* should satisfy the following equations:

$$
\begin{aligned}
p^t &= \frac{v^t \nabla f(q^t)}{f(q^t)}; \quad t = 1, \ldots, T \\
&= \left[\frac{v^t}{f(q^t)}\right] \nabla f(q^t) \\
&= P^t \nabla f(q^t)
\end{aligned} \tag{11.141}
$$

where $v^t \equiv p^t \cdot q^t$ is period t expenditure on the N commodities, P^t is the *period t aggregate price level* defined as $v^t / f(q^t) = v^t / Q^t$ and $Q^t \equiv f(q^t)$ is the corresponding *period t aggregate quantity level* for $t = 1, \ldots, T$.

Since $f(q)$ is defined by (11.140), $\nabla f(q^t) = \alpha \equiv [\alpha_1, \ldots, \alpha_N]$ for $t = 1, \ldots, T$. Substitute these equations into Eq. (11.141) and we obtain the following equations which should hold exactly under our assumptions:

$$p_{tn} = \pi_t \alpha_n; \quad n = 1, \ldots, N; \quad t = 1, \ldots, T \tag{11.142}$$

where we have redefined the period t price levels P^t in Eq. (11.141) as the parameters π_t for $t = 1, \ldots, T$.

Note that Eq. (11.142) form the basis for the *time dummy hedonic regression model*, which is due to Court (1939).[125]

At this point, it is necessary to point out that our consumer theory derivation of Eq. (11.142) is not accepted by all economists. Rosen

[124] See section "Relationships between Some Share Weighted Price Indexes" in Diewert (2021a).

[125] This was Court's (1939, 109–111) hedonic suggestion number two. He transformed the underlying Eq. (11.142) by taking logarithms of both sides of these equations (which will be done below). He chose to transform the prices by the log transformation because the resulting regression model fit his data on automobiles better. Diewert (2003b) also recommended the log transformation on the grounds that multiplicative errors were more plausible than additive errors.

(1974), Triplett (1987, 2004) and Pakes (2001)[126] have argued for a more general approach to the derivation of hedonic regression models that is based on supply conditions as well as on demand conditions. The present approach is obviously based on consumer demands and preferences only. This consumer oriented approach was endorsed by Griliches (1971, 14–15), Muellbauer (1974, 988) and Diewert (2003a, 2003b).[127] Of course, the assumption that purchasers have the same linear utility function is quite restrictive but nevertheless, it is useful to imbed hedonic regression models in a traditional consumer demand setting.

Empirically, Eq. (11.142) are unlikely to hold exactly. Thus we assume that the exact model defined by (11.142) holds only to some degree of approximation and so error terms, e_{tn}, are added to the right hand sides of Eq. (11.142). The unknown price level parameters, $\pi \equiv [\pi_1, \ldots, \pi_T]$ and quality adjustment parameters $\alpha \equiv [\alpha_1, \ldots, \alpha_N]$, can be estimated as solutions to the following (non-linear) least squares minimization problem:

$$\min_{\alpha,\pi} \left\{ \sum_{n=1}^{N} \sum_{t=1}^{T} [p_{tn} - \pi_t \alpha_n]^2 \right\}. \tag{11.143}$$

[126] "The derivatives of a hedonic price function should not be interpreted as either willingness to pay derivatives or cost derivatives; rather they are formed from a complex equilibrium process." Ariel Pakes (2001, 14).

[127] Diewert (2003b, 97) justified the consumer demand approach as follows: "After all, the purpose of the hedonic exercise is to find how demanders (and not suppliers) of the product value alternative models in a given period. Thus for the present purpose, it is the preferences of consumers that should be decisive, and not the technology and market power of producers. The situation is similar to ordinary general equilibrium theory where an equilibrium price and quantity for each commodity is determined by the interaction of consumer preferences and producer's technology sets and market power. However, there is a big branch of applied econometrics that ignores this complex interaction and simply uses information on the prices that consumers face, the quantities that they demand and perhaps demographic information in order to estimate systems of consumer demand functions. Then these estimated demand functions are used to form estimated consumer utility functions and these functions are often used in applied welfare economics. What producers are doing is entirely irrelevant to these exercises in applied econometrics with the exception of the prices that they are offering to sell at. In other words, we do not need information on producer marginal costs and markups in order to estimate consumer preferences: all we need are selling prices." Footnote 25 on page 82 of Diewert (2003b) explained how the present hedonic model can be derived from Diewert's (2003a) consumer-based model by strengthening the assumptions in the 2003a paper.

Our approach to the specification of the error terms will not be very precise. Throughout this chapter, we will obtain estimators for the aggregate price levels π_t and the quality adjustment parameters α_n as solutions to least squares minimization problems like those defined by (11.143) or as solutions to weighted least squares minimization problems that will be considered in subsequent sections. Our focus will not be on the distributional aspects of our estimators; rather, our focus will be on the *axiomatic* or *test properties* of the price levels that are solutions to the various least squares minimization problems.[128] Basically, the approach taken here is a descriptive statistics approach: we consider simple models that aggregate price and quantity information for a given period over a set of specified commodities into scalar measures of aggregate price and quantity that summarize the detailed price and quantity information in a "sensible" way.[129]

The first-order necessary (and sufficient) conditions for $\pi \equiv [\pi_1, \ldots, \pi_T]$ and $\alpha \equiv [\alpha_1, \ldots, \alpha_N]$ to solve the minimization problem defined by (11.143) are equivalent to the following $N + T$ equations:

$$\alpha_n = \frac{\sum_{t=1}^{T} \pi_t p_{tn}}{\sum_{t=1}^{T} \pi_t^2} \quad n = 1, \ldots, N$$
$$= \frac{\sum_{t=1}^{T} \pi_t^2 (p_{tn}/\pi_t)}{\sum_{t=1}^{T} \pi_t^2}; \tag{11.144}$$
$$\pi_t = \frac{\sum_{n=1}^{N} \alpha_n p_{tn}}{\sum_{n=1}^{N} \alpha_n^2} \quad t = 1, \ldots, T$$

[128] For rigorous econometric approaches to the stochastic approach to index number theory, see Rao and Hajargasht (2016) and Gorajek (2018). These papers consider many transformations of the fundamental hedonic Eq. (11.143) and many methods for constructing averages of prices.

[129] Our approach here is broadly similar to Theil's (1967, 136–137) descriptive statistics approach to index number theory.

$$= \frac{\sum_{n=1}^{N} \alpha_n^2 (p_{tn} / \alpha_n)}{\sum_{n=1}^{N} \alpha_n^2}. \tag{11.145}$$

Solutions to the two sets of equations can readily be obtained by iterating between the two sets of equations. Thus set $\alpha^{(1)} = 1_N$ (a vector of ones of dimension N) in Eq. (11.145) and calculate the resulting $\pi^{(1)} = \left[\pi_1^{(1)}, \ldots, \pi_T^{(1)}\right]$. Then substitute $\pi^{(1)}$ into the right hand sides of Eq. (11.144) to calculate $\alpha^{(2)} \equiv \left[\alpha_1^{(2)}, \ldots, \alpha_N^{(2)}\right]$. And so on until convergence is achieved.

If $\pi^* \equiv [\pi_1^*, \ldots, \pi_T^*]$ and $\alpha^* \equiv \left[\alpha_1^*, \ldots, \alpha_N^*\right]$ is a solution to (11.144) and (11.145), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$. Then $1, \pi_2^*, \ldots, \pi_T^*$ is the sequence of fixed base aggregate price levels that is generated by the least squares minimization problem defined by (11.143).

If quantity data are available, then aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \alpha^* \cdot q^t = \sum_{n=1}^{N} \alpha_n^* q_{tn}$ for $t = 1, \ldots, T$. Estimated aggregate price levels can be obtained directly from the solution to (11.143); i.e., set $P^{t*} = \pi_t^*$ for $t = 1, \ldots, T$. Alternative price levels can be *indirectly* obtained as $P^{t**} \equiv p^t \cdot q^t / Q^{t*} = p^t \cdot q^t / \alpha^* \cdot q^t$ for $t = 1, \ldots, T$. If the optimized objective function in (11.143) is 0 (so that all errors $e_{tn}^* \equiv p_{tn} - \pi_t^* \alpha_n^*$ equal 0 for $t = 1, \ldots, T$ and $n = 1, \ldots, N$), then P^{t*} will equal P^{t**} for all t. However, usually nonzero errors will occur and so a choice between the two sets of estimators must be made.[130]

From (11.144), it can be seen that α_n^*, the quality adjustment parameter for product n, is a weighted average of the T inflation adjusted prices for product n, the p_{tn} / π_t^*, where the weight for p_{tn} / π_t^* is $\pi_t^{*2} / \sum_{\tau=1}^{T} \pi_\tau^{*2}$. This means that the weight for p_{tn} / π_t^* will be very high for periods t where general inflation is high, which seems rather arbitrary.

[130] Usually, the direct estimates for the price levels will be used in hedonic regression studies or in applications of the time product dummy method; i.e., the $P^{t*} = \pi_t^*$ estimates will be used. For statistical agencies, an advantage of the direct estimates is that they can be calculated without the use of quantity information. However, later in this chapter, we will note some advantages of the indirect method if quantity information is available.

From (11.145), it can be seen that π_t^*, the period t price level (and fixed base price index), is weighted average of the N quality adjusted prices for period t, the p_{tn}/α_n^*, where the weight for p_{tn}/α_n^* is $\alpha_n^{*2}/\sum_{i=1}^N \alpha_i^{*2}$. It is a positive feature of the method that π_t^* is a weighted average of the quality adjusted prices for period t but the quadratic nature of the weights is not an attractive feature.

In addition to having unattractive weighting properties, the estimates generated by solving the least squares minimization problem (11.143) suffer from a fatal flaw: *the estimates are not invariant to changes in the units of measurement*. In order to remedy this defect, we turn to an alternative error specification.

Instead of adding approximation errors to the exact Eq. (11.142), we could append multiplicative approximation errors. Thus, the exact equations become $p_{tn} = \pi_t \alpha_n e_{tn}$ for $n = 1, \ldots, N$ and $t = 1, \ldots, T$. Upon taking logarithms of both sides of these equations, we obtain the following system of estimating equations:

$$\begin{aligned} \ln p_{tn} &= \ln \pi_t + \ln \alpha_n + \ln e_{tn}; \quad n = 1, \ldots, N; \; t = 1, \ldots, T \\ &= \rho_t + \beta_n + \varepsilon_{tn} \end{aligned} \tag{11.146}$$

where $\rho_t \equiv \ln \pi_t$ for $t = 1, \ldots, T$ and $\beta_n \equiv \ln \alpha_n$ for $n = 1, \ldots, N$. The model defined by (11.146) is the basic *Time Product Dummy regression model* with no missing observations.[131] Now choose the ρ_t and β_n to minimize the sum of squared residuals, $\sum_{n=1}^N \sum_{t=1}^T \varepsilon_{tn}^2$. Thus let $\rho \equiv [\rho_1, \ldots, \rho_T]$ and $\beta \equiv [\beta_1, \ldots, \beta_N]$ be a solution to the following least squares minimization problem:

$$\min_{\rho, \beta} \left\{ \sum_{n=1}^N \sum_{t=1}^T [\ln p_{tn} - \rho_t - \beta_n]^2 \right\}. \tag{11.147}$$

[131] In the statistics literature, this type of model is known as a fixed effects model. A generalized version of this model (with missing observations) was proposed by Summers (1973) in the international comparison context where it is known as the Country Product Dummy regression model. A weighted version of this model (with missing observations) was proposed by Aizcorbe et al. (2000).

The first-order necessary conditions for $\rho_1, \ldots, \rho_T$ and $\beta_1, \ldots, \beta_N$ to solve (11.147) are the following $T + N$ equations:

$$N\rho_t + \sum_{n=1}^{N} \beta_n = \sum_{n=1}^{N} \ln p_{tn}; \quad t = 1, \ldots, T; \tag{11.148}$$

$$\sum_{t=1}^{T} \rho_t + T\beta_n = \sum_{t=1}^{T} \ln p_{tn}; \quad n = 1, \ldots, N. \tag{11.149}$$

Replace the ρ_t and β_n in Eqs. (11.148) and (11.149) by $\ln \pi_t$ and $\ln \alpha_n$ respectively, for $t = 1, \ldots, T$ and $n = 1, \ldots, N$. After some rearrangement, the resulting equations become:

$$\pi_t = \prod_{n=1}^{N} \left(\frac{p_{tn}}{\alpha_n}\right)^{1/N}; \quad t = 1, \ldots, T; \tag{11.150}$$

$$\alpha_n = \prod_{t=1}^{T} \left(\frac{p_{tn}}{\pi_t}\right)^{1/T}; \quad n = 1, \ldots, N. \tag{11.151}$$

Thus the period t aggregate price level, π_t, is equal to the geometric average of the N quality adjusted prices for period t, $p_{t1}/\alpha_1, \ldots, p_{tN}/\alpha_N$, while the quality adjustment factor for product n, α_n, is equal to the geometric average of the T inflation adjusted prices for product n, $p_{1n}/\pi_1, \ldots, p_{Tn}/\pi_T$. These estimators look very reasonable (if quantity weights are not available).

Solutions to (11.150) and (11.151) can readily be obtained by iterating between the two sets of equations. Thus set $\alpha^{(1)} = 1_N$ (a vector of ones of dimension N) in Eq. (11.150) and calculate the resulting $\pi^{(1)} = \left[\pi_1^{(1)}, \ldots, \pi_T^{(1)}\right]$. Then substitute $\pi^{(1)}$ into the right hand sides of Eq. (11.151) to calculate $\alpha^{(2)} \equiv \left[\alpha_1^{(2)}, \ldots, \alpha_N^{(2)}\right]$. And so on until convergence is achieved. Alternatively, Eqs. (11.148) and (11.149) are linear in the unknown parameters and can be solved (after normalizing one parameter) by a simple matrix inversion. A final method of obtaining a solution to (11.148) and (11.149) is to apply a simple linear regression model to Eq. (11.146).[132]

[132] Again we require one normalization on the parameters such as $\rho_1 = 0$.

If $\pi^* \equiv [\pi_1^*, \ldots, \pi_T^*]$ and $\alpha^* \equiv [\alpha_1^*, \ldots, \alpha_N^*]$ is a solution to (11.148) and (11.149), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$ (which corresponds to $\rho_1 = 0$). Then $1, \pi_2^*, \ldots, \pi_T^*$ is the sequence of fixed base index numbers that is generated by the least squares minimization problem defined by (11.147).

Once we have the unique solution $1, \pi_2^*, \ldots, \pi_T^*$ for the T price levels that are generated by solving (11.147) along with the normalization $\pi_1 = 1$, the *price index* between period t relative to period s can be defined as π_t^*/π_s^*. Using Eq. (11.150) for π_t^* and π_s^*, we have the following expression for these price indexes:

$$\frac{\pi_t^*}{\pi_s^*} = \frac{\prod_{n=1}^{N} \left(p_{tn}/\alpha_n^*\right)^{1/N}}{\prod_{n=1}^{N} \left(p_{sn}/\alpha_n^*\right)^{1/N}} = \prod_{n=1}^{N} \left(\frac{p_{tn}}{p_{sn}}\right)^{1/N}. \tag{11.152}$$

Thus if there are no missing observations, the Time Product Dummy price indexes between any two periods in the window of T period under consideration is equal to the *Jevons index* between the two periods (the simple geometric mean of the price ratios, p_{tn}/p_{sn}).[133] This is a somewhat disappointing result since an equally weighted average of the price ratios is not necessarily a representative average of the prices; i.e., unimportant products to purchasers (in the sense that they spend very little on these products) are given the same weight in the Jevons measure of inflation between the two periods as is given to high expenditure products.[134]

Since there are no missing observations, then it can be seen using Eq. (11.151) that the ratio of the quality adjustment factor for product n

[133] This result is a special case of a more general result obtained by Triplett and McDonald (1977, 150).

[134] However, if quantity data are not available, the Jevons index has the strongest axiomatic properties; see Diewert (2021b).

relative to product m is equal to the following sensible expression:

$$\frac{\alpha_n^*}{\alpha_m^*} = \frac{\prod_{t=1}^{T} \left(p_{tn} / \pi_t^*\right)^{1/T}}{\prod_{t=1}^{T} \left(p_{tm} / \pi_t^*\right)^{1/T}} = \prod_{t=1}^{T} \left(\frac{p_{tn}}{p_{tm}}\right)^{1/T}. \tag{11.153}$$

If quantity data are available, then aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \alpha^* \cdot q^t = \sum_{n=1}^{N} \alpha_n^* q_{tn}$ for $t = 1, \ldots, T$. Estimated aggregate price levels can be obtained directly from the solution to (11.147); i.e., set $P^{t*} = \pi_t^*$ for $t = 1, \ldots, T$. Alternative price levels can be obtained *indirectly* as $P^{t**} \equiv p^t \cdot q^t / Q^{t*} = p^t \cdot q^t / \alpha^* \cdot q^t$ for $t = 1, \ldots, T$.[135] If the optimized objective function in (11.147) is 0 (so that all errors $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ equal 0 for $t = 1, \ldots, T$ and $n = 1, \ldots, N$), then P^{t*} will equal P^{t**} for all t. If the estimated residuals are not all equal to 0, then the two estimates for the period t price level P^t will differ in general. The two alternative estimates for P^t will generate different estimates for the companion aggregate quantity levels.

Note that the underlying exact model ($p_{tn} = \pi_t \alpha_n$ for all t and n) is the same for both least squares minimization problems (11.143) and (11.147). However, different error specifications and different transformations of both sides of the equations $p_{tn} = \pi_t \alpha_n$ can lead to very different estimators for the π_t and α_n. Our strategy in this section and in the following sections will be to choose specifications of the least squares minimization problem that lead to estimators for the price levels π_t that have good axiomatic properties.[136] From this perspective, it is clear that (11.147) leads to "better" estimates than (11.143).

In the following section, we allow for missing observations.

[135] The fact that a time dummy hedonic regression model generates two alternative decompositions of the value aggregate into price and quantity aggregates was first noted in de Haan and Krsinich (2018).

[136] From the perspective of the economic approach to index number theory, the minimization problems (11.143) and (11.147) have exactly the same justification; i.e., they are based on the same economic model of consumer behavior.

Time Product Dummy Regressions: The Case of Missing Observations

In this section, the least squares minimization problem defined by (11.147) is generalized to allow for missing observations. In order to make this generalization, it is first necessary to make some definitions. As in the previous section, there are N products and T time periods but not all products are purchased (or sold) in all time periods. For each period t, define the set of products n that are present in period t as $S(t) \equiv \{n : p_{tn} > 0\}$ for $t = 1, 2, \ldots, T$. It is assumed that these sets are not empty; i.e., at least one product is purchased in each period. For each product n, define the set of periods t where product n is present as $S^*(n) \equiv \{t : p_{tn} > 0\}$. Again, assume that these sets are not empty; i.e., each product is sold in at least one time period. Define the integers $N(t)$ and $T(n)$ as follows:

$$N(t) \equiv \sum_{n \in S(t)} 1; \quad t = 1, \ldots, T; \tag{11.154}$$

$$T(n) \equiv \sum_{t \in S^*(n)} 1; \quad n = 1, \ldots, N. \tag{11.155}$$

If all N products are present in period t, then $N(t) = N$; if product n is present in all T periods, then $T(n) = T$.

The multilateral methods studied in previous sections assumed that reservation prices were available for missing products in any period. Thus the methods discussed up until the present section assumed that there were no missing product prices: p_{tn} was either an actual period t price for product n or an estimated price for the product if it was missing in period t. When discussing the time product dummy multilateral price levels and indexes, we do *not* assume that reservation prices for missing products have been estimated. Instead, the method generates estimated prices for the missing products.

Using the above notation for missing products, the counterpart to (11.147) when there are missing products is the following least squares minimization problem:

$$\min_{\rho,\beta}\left\{\sum_{t=1}^{T}\sum_{n\in S(t)}[\ln p_{tn}-\rho_t-\beta_n]^2\right\}=\min_{\rho,\beta}\left\{\sum_{n=1}^{N}\sum_{t\in S^*(n)}[\ln p_{tn}-\rho_t-\beta_n]^2\right\}. \tag{11.156}$$

Note that there are two equivalent ways of writing the least squares minimization problem.[137] The first-order necessary conditions for $\rho_1, \ldots, \rho_T$ and $\beta_1, \ldots, \beta_N$ to solve (11.156) are the following counterparts to (11.148) and (11.149):

$$\sum_{n\in S(t)}[\rho_t+\beta_n]=\sum_{n\in S(t)}\ln p_{tn};\quad t=1,\ldots,T; \tag{11.157}$$

$$\sum_{t\in S^*(n)}[\rho_t+\beta_n]=\sum_{t\in S^*(n)}\ln p_{tn};\quad n=1,\ldots,N. \tag{11.158}$$

As in the previous section, let $\rho_t \equiv \ln \pi_t$ for $t = 1, \ldots, T$ and let $\beta_n \equiv \ln \alpha_n$ for $n = 1, \ldots, N$. Substitute these definitions into Eqs. (11.157) and (11.158). After some rearrangement and using definitions (11.154) and (11.155), Eqs. (11.157) and (11.158) become the following ones:

$$\pi_t=\prod_{n\in S(t)}\left[\frac{p_{tn}}{\alpha_n}\right]^{1/N(t)};\quad t=1,\ldots,T; \tag{11.159}$$

$$\alpha_n=\prod_{t\in S^*(n)}\left[\frac{p_{tn}}{\pi_t}\right]^{1/T(n)};\quad n=1,\ldots,N. \tag{11.160}$$

The same iterative procedure that was explained in the previous section will work to generate a solution to Eqs. (11.159) and (11.160).[138] As was

[137] The first expression is used when (11.156) is differentiated with respect to ρ_t and the second expression is used when differentiating (11.156) with respect to β_n.

[138] Of course, it is not necessary to use the iterative procedure to find a solution to Eqs. (11.157) and (11.158). After setting $\rho_1 = 0$ and dropping the first equation in (11.157), matrix algebra can be used to find a solution to the remaining equations. Alternatively, after setting $\rho_1 = 0$, use the equations $\ln p_{tn} = \rho_t + \beta_n + \varepsilon_{tn}$ for $t = 1, \ldots, T$ and $n \in S(t)$ to set up a linear regression model with time and product dummy variables and use a standard ordinary least squares econometric software package to obtain the

the case in the previous section, solutions to (11.159) and (11.160) are not unique; if π^*, α^* is a solution to (11.159) and (11.160), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$ (which corresponds to $\rho_1 = 0$). Then $1, \pi_2^*, \ldots, \pi_T^*$ is the sequence of (normalized) price levels that is generated by the least squares minimization problem defined by (11.156).[139] In this case, $\pi_t^* = \prod_{n\in S(t)} \left(p_{tn}/\alpha_n^*\right)^{1/N(t)}$ is the equally weighted geometric mean of all of the quality adjusted prices for the products that are available in period t or $t = 2, 3, \ldots, T$ and the quality adjustment factors are normalized so that $\pi_1^* = \prod_{n\in S(t)} \left(p_{1n}/\alpha_n\right)^{1/N(1)} = 1$. From (11.160), we can deduce that α_n^* will be larger for products that are relatively expensive and will be smaller for cheaper products.

Once we have the unique solution $1, \pi_2^*, \ldots, \pi_T^*$ for the T price levels that are generated by solving (11.156), the *price index* between period t relative to period r can be defined as π_t^*/π_r^*. Using Eqs. (11.159) and (11.160), we have the following expressions for π_t^*/π_r^* and α_n^*/α_m^* :

$$\frac{\pi_t^*}{\pi_r^*} = \frac{\prod_{n\in S(t)} \left[p_{tn}/\alpha_n^*\right]^{1/N(t)}}{\prod_{n\in S(r)} \left[p_{rn}/\alpha_n^*\right]^{1/N(r)}}; \quad 1 \le t,\ r \le T; \tag{11.161}$$

$$\frac{\alpha_n^*}{\alpha_m^*} = \frac{\prod_{t\in S^*(n)} \left[p_{tn}/\pi_t^*\right]^{1/T(n)}}{\prod_{t\in S^*(m)} \left[p_{tm}/\pi_t^*\right]^{1/T(m)}}; \quad 1 \le n,\ m \le N. \tag{11.162}$$

Note that, in general, the quality adjustment factors α_n^* do not cancel out for the indexes π_t^*/π_r^* defined by (11.161) as they did in the previous

solution $\rho_2^*, \ldots, \rho_T^*, \beta_1^*, \ldots, \beta_N^*$ to the linear regression model $\ln p_{tn} = \rho_t + \beta_n + \varepsilon_{tn}$ for $t = 1, \ldots, T$ and $n \in S(t)$. We need to assume that the X matrix for this linear regression model has full column rank.

[139] We need enough observations on products that are present so that a full rank condition is satisfied for Eqs. (11.157) and (11.158) after dropping one equation and setting $\rho_1 = 0$. If there is a rapid proliferation of new and disappearing products, then it may not be possible to invert the coefficient matrix that is associated with the modified Eqs. (11.157) and (11.158). In subsequent models with missing observations, we will assume that a similar full rank condition is satisfied.

section. However, these price indexes do have some good axiomatic properties.[140] If the set of available products is the same in periods r and t, then the quality adjustment factors do cancel and the price index for period t relative to period r is $\pi_t^*/\pi_r^* = \prod_{n\in S(t)} \left[p_{tn}/p_{rn}\right]^{1/N(t)}$, which is the Jevons index between periods r and t. Again, while this index is an excellent one if quantity information is not available, it is not satisfactory when quantity information is available due to its equal weighting of economically important and unimportant price ratios.[141]

There is another problematic property of the estimated price levels that are generated by solving the time product dummy hedonic model that is defined by (11.156): a product that is available only in one period out of the T periods has no influence on the aggregate price levels π_t^*.[142] To see this, consider Eqs. (11.157) and (11.158) and suppose that product n^* was available only in period t^*.[143] Equation n^* in the N equations in (11.158) becomes the equation: $[\rho_{t*} + \beta_{n*}] = \ln p_{t*n*}$. Thus once ρ_{t*} has been determined, β_{n*} can be defined as $\beta_{n*} \equiv \ln p_{t*n*} - \rho_{t*}$. Subtract the equation $[\rho_{t*} + \beta_{n*}] = \ln p_{t*n*}$ from equation t^* and the resulting equations in (11.157) can be written as Eq. (11.163). Dropping equation n^* in Eq. (11.158) leads to Eq. (11.164):

$$\sum_{n\in S(t),\, n\neq n^*} [\rho_t + \beta_n] = \sum_{n\in S(t),\, n\neq n^*} \ln p_{tn}; \quad t = 1, \ldots, T; \qquad (11.163)$$

$$\sum_{t\in S^*(n)} [\rho_t + \beta_n] = \sum_{t\in S^*(n)} \ln p_{tn}; \quad n = 1, \ldots, n^* - 1, n^* + 1, \ldots, N. \qquad (11.164)$$

[140] The index π_t^*/π_r^* satisfies the identity test (if prices are the same in periods r and t, then the index is equal to 1) and it is invariant to changes in the units of measurement. It is also homogeneous of degree one in the prices of period t and homogeneous of degree minus one in the prices of period r.

[141] However, if the estimated squared residuals are small in magnitude for periods τ and t, then the index π_t^*/π_r^* defined by (11.161) will be satisfactory, since in this case $p^\tau \approx \pi_t^*\alpha^*$ and $p^t \approx \pi_t^*\alpha^*$ so that prices are approximately proportional for these two periods and π_t^*/π_r^* defined by (11.161) will be approximately correct. Any missing prices for any period t and product n are defined as $p_{tn}^* \equiv \pi_t^*\alpha_n^*$.

[142] This property of the Time Product Dummy model was first noticed by Diewert (2004) (in the context of the Country Product Dummy model).

[143] We assume that products other than product n^* are available in period t^*.

Equations (11.163) and (11.164) are $T + N - 1$ equations that do not involve p_{t*n*}. After making the normalization $\rho_1^* = 0$, these equations can be solved for $\rho_2^*, \ldots, \rho_T^*, \beta_1^*, \ldots, \beta_{n*-1}^*, \beta_{n*+1}^*, \ldots, \beta_N^*$. Now define $\beta_{n*}^* \equiv \ln p_{t*n*} - \rho_{t*}$ and we have the (normalized) solution for (11.156). Since the ρ_t^* do not involve p_{t*n*}, the resulting $\pi_t^* \equiv \exp[\rho_t^*]$ for $t = 1, \ldots, T$ also do not depend on the isolated price p_{t*n*}. This proof can be repeated for any number of isolated prices. This property of the time product dummy model is unfortunate because it means that when a new product enters the marketplace in period T, it has no influence on the price levels $1, \pi_2^*, \ldots, \pi_T^*$ that are generated by solving the least squares minimization problem defined by (11.156). In other words, an expansion in the choice of products available to consumers will have no effect on price levels.

If quantity data are available, then aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \sum_{n \in S(t)} \alpha_n^* q_{tn}$ for $t = 1, \ldots, T$.[144] Estimated aggregate price levels can be obtained directly from the solution to (11.42); i.e., set $P^{t*} = \pi_t^*$ for $t = 1, \ldots, T$. Alternative price levels can be obtained indirectly as $P^{t**} \equiv \sum_{n \in S(t)} (p_n q_{tn} / Q^{t*}) = \left(\sum_{n \in S(t)} p_n q_{tn}\right) / \left(\sum_{n \in S(t)} \alpha_n^* q_{tn}\right)$ for $t = 1, \ldots, T$.[145] If the optimized objective function in (11.156) is 0, so that all errors $\varepsilon_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ equal 0 for $t = 1, \ldots, T$ and $n \in S(t)$, then P^{t*} will equal P^{t**} for all t. If the estimated residuals are not all equal to 0, then the two estimates

[144] Note that each $\alpha_n^* > 0$ since $\alpha_n^* \equiv \exp[\beta_n^*]$ for $n = 1, \ldots, N$.

[145] Note that $P^{t**} \equiv \left(\sum_{n \in S(t)} p_{tn} q_{tn}\right) / \left(\sum_{n \in S(t)} \alpha_n^* q_{tn}\right)$ is a period t *quality adjusted unit value price level*; see section "Quality Adjusted Unit Value Price and Quantity Indexes". The corresponding quantity level is $Q^{t**} \equiv \left(\sum_{n \in S(t)} p_{tn} q_{tn}\right) / P^{t**} = \sum_{n \in S(t)} \alpha_n^* q_{tn}$, which is the level generated by a *linear aggregator function*. By looking at (11.156), it can be seen that if prices are identical in periods t and r so that $p^t = p^r$, then $P^{t*} = P^{r*}$; i.e., an identity test for the direct hedonic price levels will be satisfied. However, the corresponding Q^{t*} will not satisfy the identity test for quantity levels; i.e., if quantities q_{tn} and q_{rn} are equal in periods t and r for all n, it is not the case that $Q^{t*} \equiv \sum_{n=1}^N (p_{tn} q_{tn} / \pi_t^*)$ will equal $Q^{r*} \equiv \sum_{n=1}^N (p_{rn} q_{rn} / \pi_r^*)$ for $r \neq t$ unless prices are also equal for the two periods. On the other hand, it can be seen that $Q^{t**} = \sum_{n \in S(t)} \alpha_n^* q_{tn} = \sum_{n \in S(t)} \alpha_n^* q_{rn} = Q^{r*}$ if $q_{tn} = q_{rn}$ for all n even if prices are not identical for the two periods. Thus the choice between using P^{t*} or P^{t**} could be made on the basis of choosing which identity test is more important to satisfy. The analysis here follows that of de Haan and Krsinich (2018, 763–764).

for the period t price level P^t will differ. The two estimates for P^t will generate different estimates for the companion aggregate quantity levels.

Weighted Time Product Dummy Regressions: The Bilateral Case

A major problem with the indexes discussed in the previous two sections is the fact that they do not weight the individual product prices by their economic importance. The first serious index number economist to stress the importance of weighting was Walsh (1901).[146] Keynes was quick to follow up on the importance of weighting[147] and Fisher emphatically endorsed weighting.[148] Griliches also endorsed weighting in the hedonic regression context.[149]

[146] See Walsh (1901). This book laid the groundwork for the test or axiomatic approach to index number theory that was further developed by Fisher (1922). In his second book on index number theory, Walsh made the case for weighting by economic importance as follows: "It might seem at first sight as if simply every price quotation were a single item, and since every commodity (any kind of commodity) has one price-quotation attached to it, it would seem as if price-variations of every kind of commodity were the single item in question. This is the way the question struck the first inquirers into price-variations, wherefore they used simple averaging with even weighting. But a price-quotation is the quotation of the price of a generic name for many articles; and one such generic name covers a few articles, and another covers many. … A single price-quotation, therefore, may be the quotation of the price of a hundred, a thousand, or a million dollar's worths, of the articles that make up the commodity named. Its weight in the averaging, therefore, ought to be according to these money-unit's worth." Correa Moylan Walsh (1921a, 82–83).

[147] "It is also clear that the so-called unweighted index numbers, usually employed by practical statisticians, are the worst of all and are liable to large errors which could have been easily avoided." J. M. Keynes (1909, 79). This paper won the Cambridge University Adam Smith Prize for that year. Keynes (1930, 76–77) again stressed the importance of weighting in a later paper which drew heavily on his 1909 paper.

[148] "It has already been observed that the purpose of any index number is to strike a fair average of the price movements or movements of other groups of magnitudes. At first a simple average seemed fair, just because it treated all terms alike. And, in the absence of any knowledge of the relative importance of the various commodities included in the average, the simple average is fair. But it was early recognized that there are enormous differences in importance. Everyone knows that pork is more important than coffee and wheat than quinine. Thus the quest for fairness led to the introduction of weighting." Irving Fisher (1922, 43).

[149] "But even here, we should use a weighted regression approach, since we are interested in an estimate of a weighted average of the pure price change, rather than just

In this section, we will discuss some alternative methods for weighting by economic importance in the context of a bilateral time product dummy regression model.[150] We also assume that there are no missing observations in this section.

Recall the least squares minimization problem defined by (11.147) in section "Time Product Dummy Regressions: The Case of No Missing Observations". The squared residuals $[\ln p_{tn} - \rho_t - \beta_n]^2$, appear in this problem without any weighting. Thus products, which have a high volume of sales in any period, are given the same weight in the least squares minimization problem as products that have very few sales. In order to take economic importance into account, for the case of two time periods, replace (11.147) by the following *weighted least squares minimization problem*:

$$\min_{\rho,\beta}\left\{\sum_{n=1}^{N} q_{1n}[\ln p_{1n} - \beta_n]^2 + \sum_{n=1}^{N} q_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2\right\} \quad (11.165)$$

where we have set $\rho_1 = 0$. The squared error for product n in period t is repeated q_{tn} times to reflect the sales of the product in period t. Thus the new problem (11.165) takes into account the popularity of each product.[151]

The first-order necessary conditions for the minimization problem defined by (11.165) are the following $N + 1$ equations:

$$(q_{1n} + q_{2n})\beta_n = q_{1n} \ln p_{1n} + q_{2n}(\ln p_{2n} - \rho_2); \quad n = 1, \ldots, N; \quad (11.166)$$

an unweighted average over all possible models, no matter how peculiar or rare." Zvi Griliches (1971, 8).

[150] The approach taken in this section is based on Rao (1995, 2004, 2005) and Diewert (2003b, 2005a, 2005b). Diewert (2005a) considered all four forms of weighting that will be discussed in this section while Rao (1995, 2005) discussed mainly the third form of weighting.

[151] One can think of repeating the term $[\ln p_{1n} - \beta_n]^2$ for each unit of product n sold in period 1. The result is the term $q_{1n}[\ln p_{1n} - \beta_n]^2$. A similar justification based on repeating the price according to its sales can also be made. This repetition methodology makes the stochastic specification of the error terms somewhat complicated. However, as indicated in the introduction, we leave these difficult distributional problems to other more capable econometricians.

$$\left(\sum_{n=1}^{N} q_{2n}\right)\rho_2 = \sum_{n=1}^{N} q_{2n}(\ln p_{2n} - \beta_n). \tag{11.167}$$

The solution to (11.166) and (11.167) is the following one[152]:

$$\rho_2^* \equiv \frac{\sum_{n=1}^{N} q_{1n}q_{2n}(q_{1n}+q_{2n})^{-1}\ln\left(p_{2n}/p_{1n}\right)}{\sum_{i=1}^{N} q_{1i}q_{2i}(q_{1i}+q_{2i})^{-1}}; \tag{11.168}$$

$$\beta_n^* \equiv q_{1n}(q_{1n}+q_{2n})^{-1}\ln(p_{1n}) + q_{2n}(q_{1n}+q_{2n})^{-1}\ln\left(\frac{p_{2n}}{\pi_2^*}\right); \quad n = 1, \ldots, N \tag{11.169}$$

where $\pi_2^* \equiv \exp\left[\rho_2^*\right]$. Note that the weight for the term $\ln\left(p_{2n}/p_{1n}\right)$ in (11.168) can be written as follows:

$$\begin{aligned} q_n^* &\equiv \frac{\sum_{n=1}^{N} q_{1n}q_{2n}(q_{1n}+q_{2n})^{-1}}{\sum_{i=1}^{N} q_{1i}q_{2i}(q_{1i}+q_{2i})^{-1}}; \quad n = 1, \ldots, N \\ &= \frac{h(q_{1n}, q_{2n})}{\sum_{i=1}^{N} h(q_{1i}, q_{2i})} \end{aligned} \tag{11.170}$$

where $h(a, b) \equiv 2ab/(a+b) = \left[(1/2)a^{-1} + (1/2)b^{-1}\right]^{-1}$ is the *harmonic mean* of a and b.[153]

Note that the q_n^* sum to 1 and thus ρ_2^* is a weighted average of the logarithmic price ratios $\ln\left(p_{2n}/p_{1n}\right)$. Using $\pi_2^* = \exp\left[\rho_2^*\right]$ and $\pi_1^* = \exp\left[\rho_1^*\right] = \exp[0] = 1$, the bilateral price index that is generated by the

[152] See Diewert (2005a).

[153] $h(a, b)$ is well defined by $ab/(a+b)$ if a and b are nonnegative and at least one of these numbers is positive. In order to write $h(a, b)$ as $\left[(1/2)a^{-1} + (1/2)b^{-1}\right]^{-1}$, we require $a > 0$ and $b > 0$.

solution to (11.165) is

$$\frac{\pi_2^*}{\pi_1^*} = \exp[\rho_2^*] = \exp\left[\sum_{n=1}^{N} q_n^* \ln\left(\frac{p_{2n}}{p_{1n}}\right)\right]. \quad (11.171)$$

Thus π_2^*/π_1^* is a weighted geometric mean of the price ratios p_{2n}/p_{1n} with weights q_n^* defined by (11.170). Although this seems to be a reasonable bilateral index number formula, it must be rejected for practical use on the grounds that *the index is not invariant to changes in the units of measurement*.

Since values are invariant to changes in the units of measurement, the lack of invariance problem can be solved if we replace the quantity weights in (11.165) with expenditure or sales weights.[154] This leads to the following weighted least squares minimization problem where the weights v_{tn} are defined as $p_{tn}q_{tn}$ for $t = 1, 2$ and $n = 1, \ldots, N$:

$$\min_{\rho,\beta}\left\{\sum_{n=1}^{N} v_{1n}[\ln p_{1n} - \beta_n]^2 + \sum_{n=1}^{N} v_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2\right\}. \quad (11.172)$$

It can be seen that problem (11.172) has exactly the same mathematical form as problem (11.165) except that v_{tn} has replaced q_{tn} and so the solutions (11.168) and (11.169) will be valid in the present context if v_{tn} replaces q_{tn} in these formulae. Thus the solution to (11.172) is:

$$\rho_2^* \equiv \frac{\sum_{n=1}^{N} v_{1n}v_{2n}(v_{1n} + v_{2n})^{-1}\ln(p_{2n}/p_{1n})}{\sum_{i=1}^{N} v_{1i}v_{2i}(v_{1i} + v_{2i})^{-1}}; \quad (11.173)$$

$$\beta_n^* \equiv v_{1n}(v_{1n} + v_{2n})^{-1}\ln(p_{1n})$$

154 "But on what principle shall we weight the terms? Arthur Young's guess and other guesses at weighting represent, consciously or unconsciously, the idea that relative money values of the various commodities should determine their weights. A value is, of course, the product of a price per unit, multiplied by the number of units taken. Such values afford the only common measure for comparing the streams of commodities produced, exchanged, or consumed, and afford almost the only basis of weighting which has ever been seriously proposed." Irving Fisher (1922, 45).

$$+ v_{2n}(v_{1n} + v_{2n})^{-1} \ln\left(\frac{p_{2n}}{\pi_2^*}\right); \quad n = 1, \ldots, N \tag{11.174}$$

where $\pi_2^* \equiv \exp[\rho_2^*]$.

The resulting price index, $\pi_2^*/\pi_1^* = \pi_2^* = \exp[\rho_2^*]$ is indeed invariant to changes in the units of measurement. However, if we regard π_2^* as a function of the price and quantity vectors for the two periods, say $P(p^1, p^2, q^1, q^2)$, then another problem emerges for the price index defined by the solution to (11.172): $P(p^1, p^2, q^1, q^2)$ is not homogeneous of degree 0 in the components of q^1 or in the components of q^2. These properties are important because it is desirable that the companion implicit quantity index defined as $Q(p^1, p^2, q^1, q^2) \equiv (p^2 \cdot q^2/p^1 \cdot q^1)/P(p^1, p^2, q^1, q^2)$ be homogeneous of degree 1 in the components of q^2 and homogeneous of degree minus 1 in the components of q^1.[155] We also want $P(p^1, p^2, q^1, q^2)$ to be homogeneous of degree 1 in the components of p^2 and homogeneous of degree minus 1 in the components of p^1 and these properties are also not satisfied. Thus, we conclude that the solution to the weighted least squares problem defined by (11.172) does not generate a satisfactory price index formula.

The above deficiencies can be remedied if the *expenditure amounts* v_{tn} in (11.172) are replaced by *expenditure shares*, s_{tn}. Where $v_t \equiv \sum_{n=1}^{N} v_{tn}$ for $t = 1, 2$ and $s_{tn} \equiv v_{tn}/v_t$ for $t = 1, 2$ and $n = 1, \ldots, N$. This replacement leads to the following weighted least squares minimization

[155] Thus, we want Q to have the following properties: $Q(p^1, p^2, q^1, \lambda q^2) = \lambda Q(p^1, p^2, q^1, q^2)$ and $Q(p^1, p^2, \lambda q^1, q^2) = \lambda^{-1} Q(p^1, p^2, q^1, q^2)$ for all $\lambda > 0$.

problem[156]:

$$\min_{\rho,\beta}\left\{\sum_{n=1}^{N} s_{1n}[\ln p_{1n}-\beta_n]^2+\sum_{n=1}^{N} s_{2n}[\ln p_{2n}-\rho_2-\beta_n]^2\right\}. \tag{11.175}$$

Again, it can be seen that problem (11.175) has exactly the same mathematical form as problem (11.165) except that s_{tn} has replaced q_{tn} and so the solutions (11.168) and (11.169) will be valid in the present context if s_{tn} replaces q_{tn} in these formulae. Thus, the solution to (11.175) is:

$$\rho_2^* \equiv \frac{\sum_{n=1}^{N} s_{1n}s_{2n}(s_{1n}+s_{2n})^{-1}\ln(p_{2n}/p_{1n})}{\sum_{i=1}^{N} s_{1i}s_{2i}(s_{1i}+s_{2i})^{-1}}; \tag{11.176}$$

$$\beta_n^* \equiv s_{1n}(s_{1n}+s_{2n})^{-1}\ln p_{1n} + s_{2n}(s_{1n}+s_{2n})^{-1}\ln\left(\frac{p_{2n}}{\pi_2^*}\right); \quad n=1,\ldots,N \tag{11.177}$$

where $\pi_2^* \equiv \exp[\rho_2^*]$. Define the *normalized harmonic mean share weights* as $s_n^* \equiv h(s_{1n}, s_{2n})\Big/\sum_{i=1}^{N} h(s_{1i}, s_{2i})$ for $n = 1, \ldots, N$. Then the weighted time product dummy bilateral price index, $P_{\mathrm{WTPD}}(p^1, p^2, q^1, q^2) \equiv \pi_2^*/\pi_1^* = \pi_2^*$, has the following logarithm:

$$\ln P_{\mathrm{WTPD}}\left(p^1, p^2, q^1, q^2\right) \equiv \sum_{n=1}^{N} s_n^* \ln\left(\frac{p_{2n}}{p_{1n}}\right). \tag{11.178}$$

[156] Note that the minimization problem defined by (11.175) is equivalent to the problem of minimizing $\sum_{n=1}^{N} e_{1n}^2 + \sum_{n=1}^{N} e_{2n}^2$ with respect to $\rho_2, \beta_1, \ldots, \beta_N$ where the error terms e_{tn} are defined by the equations $s_{1n}^{1/2}\ln p_{1n} = s_{1n}^{1/2}\beta_n + e_{1n}$ for $n = 1, \ldots, N$ and $s_{2n}^{1/2}\ln p_{2n} = s_{2n}^{1/2}\rho_2 + s_{2n}^{1/2}\beta_n + e_{2n}$ for $n = 1, \ldots, N$. Thus the solution to (11.175) can be found by running a linear regression using the above two sets of estimating equations. The numerical equivalence of the least squares estimates obtained by repeating multiple observations or by using the square root of the weight transformation was noticed long ago as the following quotation indicates: "It is evident that an observation of weight w enters into the equations exactly as if it were w separate observations each of weight unity. The best practical method of accounting for the weight is, however, to prepare the equations of condition by multiplying each equation throughout by the square root of its weight." E. T. Whittaker and G. Robinson (1940, 224).

Thus $P_{WTPD}(p^1, p^2, q^1, q^2)$ is equal to a share weighted geometric mean of the price ratios, p_{2n}/p_{1n}.[157] This index is a satisfactory one from the viewpoint of the test approach to index number theory. It can be shown that $P_{WTPD}(p^1, p^2, q^1, q^2)$ satisfies the following tests:

(i) the *identity test*; i.e., $P_{WTPD}(p^1, p^2, q^1, q^2) = 1$ if $p^1 = p^2$;
(ii) the *time reversal test*; i.e.,$P_{WTPD}(p^2, p^1, q^2, q^1) = 1/P_{WTPD}(p^1, p^2, q^1, q^2)$;[158]
(iii) *homogeneity of degree 1 in period 2 prices*; i.e., $P_{WTPD}(p^1, \lambda p^2, q^1, q^2) = \lambda P_{WTPD}(p^1, p^2, q^1, q^2)$;
(iv) *homogeneity of degree −1 in period 1 prices*; i.e.,

$$P_{WTPD}\left(\lambda p^1, p^2, q^1, q^2\right) = \lambda^{-1} P_{WTPD}\left(p^1, p^2, q^1, q^2\right);$$

(v) *homogeneity of degree 0 in period 1 quantities*; i.e.,

$$P_{WTPD}\left(p^1, p^2, \lambda q^1, q^2\right) = P_{WTPD}\left(p^1, p^2, q^1, q^2\right);$$

(vi) *homogeneity of degree 0 in period 2 quantities*; i.e.,

$$P_{WTPD}\left(p^1, p^2, q^1, \lambda q^2\right) = P_{WTPD}\left(p^1, p^2, q^1, q^2\right);$$

(vii) *invariance to changes in the units of measurement*;
(viii) the *min–max test*; i.e., $\min_n\{p_{2n}/p_{1n} : n = 1, \ldots, N\} \leq P_{WTPD}(p^1, p^2, q^1, q^2) \leq \max_n\{p_{2n}/p_{1n} : n = 1, \ldots, N\}$; and
(ix) the *invariance to the ordering of the products test*.

Moreover, it can be shown that $P_{WTPD}(p^1, p^2, q^1, q^2)$ approximates the superlative Törnqvist Theil index to the second order around an equal price and quantity point where $p^1 = p^2$ and $q^1 = q^2$.[159] Thus if changes in prices and quantities going from one period to the next are not too

[157] See Diewert (2002, 2005a).

[158] See Diewert (2003b, 2005b).

[159] Diewert (2005a, 564) noted this result. Thus P_{WTPD} is a pseudo-superlative index. For the definition of a superlative index, see Diewert (1976, 2021a). A pseudo-superlative index approximates a superlative index to the second order around any point where $p^1 = p^2$ and $q^1 = q^2$; see Diewert (1978).

large and there are no missing products, P_{WTPD} should be close to the superlative Fisher (1922) and Törnqvist Theil indexes.[160]

Recall the results from section "Time Product Dummy Regressions: The Case of No Missing Observations" for the unweighted time product dummy model. From Eq. (11.152), it can be seen that the unweighted bilateral time product dummy regression model generated the Jevons index as the solution to the unweighted least squares minimization problem that is a counterpart to the weighted problem defined by (11.175) above. Thus, appropriate weighting of the squared errors has changed the solution index dramatically: the index defined by (11.178) weights products by their economic importance and has good test properties whereas the Jevons index can generate very problematic results due to its lack of weighting according to economic importance. Note that both models have the same underlying structure; i.e., they assume that p_{tn} is approximately equal to $\pi_t\alpha_n$ for $t = 1, 2$ and $n = 1, \ldots, N$. *Thus weighting by economic importance has converted a least squares minimization problem that generates a rather poor price index into a problem that generates a rather good index.*

There is one more weighting scheme that generates an even better index in the bilateral context where we are running a time product dummy hedonic regression using the price and quantity data for only two periods. Consider the following weighted least squares minimization problem:

$$\min_{\rho,\beta}\left\{\sum_{n=1}^{N}\left(\frac{1}{2}\right)(s_{1n}+s_{2n})[\ln p_{1n}-\beta_n]^2+\sum_{n=1}^{N}\left(\frac{1}{2}\right)(s_{1n}+s_{2n})[\ln p_{2n}-\rho_2-\beta_n]^2\right\}. \tag{11.179}$$

As usual, it can be seen that problem (11.179) has exactly the same mathematical form as problem (11.165) except that $(1/2)(s_{1n}+s_{2n})$ has replaced q_{tn} and so the solutions (11.168) and (11.169) will be valid in the present context if $(1/2)(s_{1n}+s_{2n})$ replaces q_{tn} in these formulae.

[160] However, with large changes in price and quantities going from period 1 to 2, P_{WTPD} will tend to lie below its superlative counterparts; see Diewert (2018, 53) and an example in Diewert and Fox (2020).

Thus, the solution to (11.179) simplifies to the following solution:

$$\rho_2^* \equiv \sum_{n=1}^{N} \left(\frac{1}{2}\right)(s_{1n} + s_{2n}) \ln\left(\frac{p_{2n}}{p_{1n}}\right); \tag{11.180}$$

$$\beta_n^* \equiv \left(\frac{1}{2}\right) \ln p_{1n} + \left(\frac{1}{2}\right) \ln\left(\frac{p_{2n}}{\pi_2^*}\right); \quad n = 1, \ldots, N \tag{11.181}$$

where $\pi_2^* \equiv \exp[\rho_2^*]$ and $\pi_1^* \equiv \exp[\rho_1^*] = \exp[0] = 1$ since we have set $\rho_1{}^* = 0$. Thus, the bilateral index number formula which emerges from the solution to (11.179) is $\pi_2^*/\pi_1^* = \exp\left[\sum_{n=1}^{N} (1/2)(s_{1n} + s_{2n}) \ln(p_{2n}/p_{1n})\right] \equiv P_T(p^1, p^2, q^1, q^2)$, which is the Törnqvist Theil (1967, 137–138) bilateral index number formula. Thus, the use of the weights in (11.179) has generated an even better bilateral index number formula than the formula that resulted from the use of the weights in (11.175). This result reinforces the case for using appropriately weighted versions of the basic time product dummy hedonic regression model.[161] However, if the implied residuals in the original unweighted minimization problem (11.147) are small (or equivalently, if the fit in the linear regression model that can be associated with (11.147) is high so that predicted values for log prices are close to actual log prices), then *weighting will not matter very much* and the unweighted model [11.147] will give results that are similar to the results generated by the weighted model defined by (11.179). This comment applies to all of the weighted hedonic regression models that are considered in this paper.[162]

The aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \alpha^* \cdot q^t = \sum_{n=1}^{N} \alpha_n^* q_{tn}$ for $t = 1, 2$ where the α_n^* are defined as the exponentials of the β_n^* defined by (11.181). Estimated aggregate price levels can be obtained directly from the solution to (11.179); i.e.,

[161] Note that the bilateral regression model defined by the minimization problem (11.175) is readily generalized to the case of T periods whereas the bilateral regression model defined by the minimization problem (11.179) cannot be generalized to the case of T periods. These facts were noted by de Haan and Krsinich (2014).

[162] If the residuals are small for (11.147), then prices will vary almost proportionally over time and all reasonable index number formulae will register price levels that are close to the estimated π_t^*; i.e., we will have $p^t \approx \pi_t^* p^1$ for $t = 2, 3, \ldots, T$ if the residuals are small for (11.147).

set $P^{t*} = \pi_t^*$ for $t = 1, 2$.[163] Alternative price levels can be obtained indirectly as $P^{t**} \equiv p^t \cdot q^t / Q^{t*} = p^t \cdot q^t / \alpha^* \cdot q^t$ for $t = 1, 2$. If the optimized objective function in (11.179) is 0, so that all errors equal 0, then P^{t*} will equal P^{t**} for $t = 1, 2$. If the estimated residuals are not all equal to 0, then the two estimates for the period t price level P^t will differ and the alternative estimates for P^t will generate different estimates for the companion aggregate quantity levels.

It should be noted that we have not made any bias corrections due to the fact that our model estimates the logarithm of π_t instead of π_t itself. This is due to our perspective that simply tries to fit an exact model by transforming it in a way that leads to solutions π_t^* to a least squares minimization problem where the π_t^* have good axiomatic properties.[164] There is more work to be done in working out the distributional properties of the above estimators for the price levels.

Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations

In this section, we will generalize the last two models in the previous section to cover the case where there are missing observations.[165] Thus

[163] In this case, alternative period t quantity levels are defined as $Q^{1**} \equiv p^1 \cdot q^1$ and $Q^{2**} \equiv p^2 \cdot q^2 / \pi_2^* = [v_2/v_1] / P_T(p^1, p^2, q^1, q^2)$. If the squared errors in (11.179) are all 0, then the alternative quantity estimates are equal to each other and the model $\ln p_{tn} = \rho_t + \beta_n$ holds exactly for each t and n, which means that prices are proportional across the two periods; i.e., we have $p^t = \pi_t^* \alpha^*$ for $t = 1, 2$ where $\alpha^* \equiv [\alpha_1^*, \ldots, \alpha_N^*]$. In the case where the squared errors are nonzero, the π_t^*, Q^{t**} aggregates are preferred since $P_T(p^1, p^2, q^1, q^2)$ is a superlative index and thus has a strong economic justification.

[164] We note that de Haan and Krsinich (2018, 769–770) make the following comments on possible biases that result from the use of a weighted least squares model to generate price indexes: "Finally, we will elaborate on a few econometric issues. The estimated quality adjusted prices … are biased as taking exponentials is a non-linear transformation. The time dummy index is similarly biased. It is questionable whether bias adjustments would be appropriate, though, at least from an index number point of view. For instance, recall the two-period case with only matched items, where Diewert's (2004) choice of regression weights ensures that the time dummy index is equal to the superlative Törnqvist price index. Correcting for the "bias" would mean that this useful property does no longer hold, and so there is a tension between econometrics and index number theory."

[165] The results in this section are closely related to the results derived by de Haan (2004a), Silver and Heravi (2005) and de Haan and Krsinich (2014, 2018). However, our method of derivation is somewhat different.

we assume that there are products that are missing in period 2 that were present in period 1 and some new products that appear in period 2. As in section "Time Product Dummy Regressions: The Case of Missing Observations", $S(t)$ denotes the set of products n that are present in period t for $t = 1, 2$. It is assumed that $S(1) \cap S(2)$ is not the empty set; i.e., there are one or more products that are present in both periods. We need some new notation to deal with missing prices and quantities. For the present, if product n is not present in period t, define p_{tn} and q_{tn} to equal 0. This enables us to define the N dimensional period t price and quantity vectors as $p^t \equiv [p_{t1}, \ldots, p_{tN}]$ and $q^t \equiv [q_{t1}, \ldots, q_{tN}]$ for $1 = 1, 2$. Thus the missing prices and quantities are simply set equal to 0. The period t share of sales or expenditures for product n is defined in the usual case as $s_{tn} \equiv p_{tn}q_{tn} / p^t \cdot q^t$ for $n = 1, \ldots, N$ and $t = 1, 2$. With these notational conventions, the new weighted least squares minimization problem that generalizes (11.175) is the following minimization problem[166]:

$$\min_{\rho,\beta} \left\{ \sum_{n \in S(1)} s_{1n}[\ln p_{1n} - \beta_n]^2 + \sum_{n \in S(2)} s_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2 \right\}. \qquad (11.182)$$

The first-order conditions for $\rho_2^*, \beta_1^*, \ldots, \beta_N^*$ to solve (11.182) are equivalent to the following equations:

$$\sum_{n \in S(2)} s_{2n}\rho_2^* + \sum_{n \in S(2)} s_{2n}\beta_n^* = \sum_{n \in S(2)} s_{2n} \ln p_{2n}; \qquad (11.183)$$

$$s_{2n}\rho_2^* + (s_{1n} + s_{2n})\beta_n^* = s_{1n} \ln p_{1n} + s_{2n} \ln p_{1n}; \quad n \in S(1) \cap S(2); \qquad (11.184)$$

$$\beta_n^* = \ln p_{1n}; \quad n \in S(1), n \notin S(2); \qquad (11.185)$$

$$\rho_2^* + \beta_n^* = \ln p_{2n}; \quad n \in S(2), n \notin S(1). \qquad (11.186)$$

Define the intersection set of products S^* as follows:

$$S^* \equiv S(1) \cap S(2). \qquad (11.187)$$

[166] This form of weighting was suggested by Rao (1995, 2004, 2005), Diewert (2002, 2004, 2005a) and de Haan (2004a).

Substituting Eq. (11.186) into Eq. (11.183) leads to the following equation:

$$\sum_{n \in S^*} s_{2n}\left[\ln p_{2n} - \rho_2^* - \beta_n^*\right] = 0. \tag{11.188}$$

Consider the following least squares minimization problem that is defined over the set of products that are present in both periods:

$$\min_{\rho, \beta}\left\{\sum_{n \in S^*} s_{1n}[\ln p_{1n} - \beta_n]^2 + \sum_{n \in S^*} s_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2\right\}. \tag{11.189}$$

The first-order conditions for this problem are (11.188) and (11.184). Once we find the solution to this problem, define β_n^* for the products that are not present in both periods by Eqs. (11.185) and (11.186). This augmented solution will solve problem (11.182). The solution to (11.189) can be found by adapting the solution to (11.175) to the current situation. Recall Eqs. (11.176) and (11.177) from the previous section. Replacing the entire set of product indices $n = 1, \ldots, N$ by the intersection set S^* defined by (11.187) leads to the following solution to (11.189):

$$\rho_2^* \equiv \frac{\sum_{n \in S^*} s_{1n} s_{2n} (s_{1n} + s_{2n})^{-1} \ln\left(p_{2n} / p_{1n}\right)}{\sum_{i \in S^*} s_{1i} s_{2i} (s_{1i} + s_{2i})^{-1}}; \tag{11.190}$$

$$\beta_n^* \equiv s_{1n}(s_{1n} + s_{2n})^{-1} \ln p_{1n} + s_{2n}(s_{1n} + s_{2n})^{-1} \ln\left(p_{2n} / \pi_2^*\right); \quad n \in S^* \tag{11.191}$$

where $\pi_2^* \equiv \exp\left[\rho_2^*\right]$. Define the *normalized harmonic mean share weights* for the always present products as follows as $s_n^* \equiv h(s_{1n}, s_{2n}) / \sum_{i \in S^*} h(s_{1i}, s_{2i})$ for $n \in S^*$. Using these definitions for the shares s_n^*, the *weighted time product dummy bilateral price index with missing observations*, $P_{\text{WTPD}}(p^1, p^2, q^1, q^2) \equiv \pi_2^* / \pi_1^* = \pi_2^*$, has the following logarithm:

$$\ln P_{\text{WTPD}}\left(p^1, p^2, q^1, q^2\right) \equiv \sum_{n \in S^*} s_n^* \ln\left(\frac{p_{2n}}{p_{1n}}\right). \tag{11.192}$$

Note that $P_{WTPD} \equiv \pi_2^*/\pi_1^*$ depends directly on the price ratios for the products *that are present in both periods*. However, it also depends on the shares s_{tn}, which in turn depend on all of the price and quantity information for both periods. It can be seen that $P_{WTPD}(p^1, p^2, q^1, q^2)$ is a weighted geometric mean of the matched prices p_{2n}/p_{1n} for products n that are present in both periods. Thus if matched product prices are equal in the two periods, then $P_{WTPD}(p^1, p^2, q^1, q^2)$ will equal unity even if there is an expanding or contracting choice set over the two periods; i.e., alternative reservation prices for any missing products will not affect the estimated price levels and price indexes.

However, the hedonic regression model that is generated by solving (11.189) can be used to impute (neutral) reservation prices for missing observations. Thus define $\alpha_n^* \equiv \exp[\beta_n^*]$ for $n = 1, \ldots, N$. Then the missing prices p_{tn}^* can be defined as follows:

$$p_{2n}^* \equiv \pi_2^*\alpha_n^* = \pi_2^* p_{1n} \quad n \in S(1), \quad n \notin S(2); \tag{11.193}$$

$$p_{1n}^* \equiv \pi_1^*\alpha_n^* = \frac{p_{2n}}{\pi_2^*} \quad n \in S(2), \quad n \notin S(1). \tag{11.194}$$

Thus the missing prices for period 2, p_{2n}^*, are the corresponding *inflation adjusted carry forward prices* from period 1, p_{1n} times π_2^* and the missing prices for period 1, p_{1n}^*, are the corresponding *inflation adjusted carry backward prices* from period 2, p_{2n} deflated by π_2^*, where π_2^* is the weighted time product dummy price index $P_{WTPD}(p^1, p^2, q^1, q^2)$ defined as $\pi_2^* \equiv \exp[\rho_2^*]$ where ρ_2^* is defined by (11.190).[167] As noted above, these reservation prices are neutral in the sense that they do not affect the definition of ρ_2^* and hence they do not affect the definition of $P_{WTPD}(p^1, p^2, q^1, q^2)$..

Estimated aggregate price levels can be obtained directly from the solution to (11.189); i.e., set $P^{1*} = 1$ and $P^{2*} = \pi_2^*$. The corresponding quantity levels are defined as $Q^{1*} \equiv p^1 \cdot q^1$ and $Q^{2*} \equiv p^2 \cdot q^2/\pi_2^*$. Alternative price and quantity levels can be obtained as $Q^{t**} \equiv \alpha^* \cdot q^t$ and $P^{t**} \equiv p^t \cdot q^t/Q^{t**}$ for $t = 1, 2$. If the optimized objective function in (11.189) is 0, so that all errors equal 0, then P^{t*} will equal P^{t**} for all t. If the estimated residuals are not all equal to 0, then the two

[167] The corresponding imputed values for the missing quantities in each period are set equal to 0.

estimates for the period 2 price level P^2 will differ and, as usual, the alternative estimates for P^2 will generate different estimates for the companion aggregate quantity levels.

The above analysis is not quite the end of the story. The expenditure shares s_{1n} and s_{2n} which appear in (11.182) are not the expenditure shares that characterize the always present products; they are the original expenditure shares defined over all N products. It is of interest to compare $P_{\text{WTPD}}(p^1, p^2, q^1, q^2)$ defined implicitly by (11.192) with the weighted time product dummy index, $P_{\text{WTPDM}}(p^{1*}, p^{2*}, q^{1*}, q^{2*})$, that is defined over the common set of products, S^*[168]; i.e., P_{WTPDM} is the weighted time product dummy regression model that is defined over the set of *matched products* for the two periods under consideration.

Define $v_t^* \equiv \sum_{n \in S^*} v_{tn}$ as the total expenditure on always present products for $t = 1, 2$ and define the corresponding *restricted expenditure shares* as[169]:

$$s_{tn}^* \equiv \frac{v_{tn}}{v_t^*}; \quad t = 1, 2; \quad n \in S^*. \tag{11.195}$$

The matched model version of (11.189) is the following weighted least squares minimization problem:

$$\min_{\rho, \beta} \left\{ \sum_{n \in S^*} s_{1n}^* [\ln p_{1n} - \beta_n]^2 + \sum_{n \in S^*} s_{2n}^* [\ln p_{2n} - \rho_2 - \beta_n]^2 \right\}. \tag{11.196}$$

The ρ_2 solution to (11.196) is the following one:

$$\rho_2^{**} \equiv \frac{\sum_{n \in S^*} s_{1n}^* s_{2n}^* \left(s_{1n}^* + s_{2n}^*\right)^{-1} \ln\left(p_{2n} / p_{1n}\right)}{\sum_{i \in S^*} s_{1i}^* s_{2i}^* \left(s_{1i}^* + s_{2i}^*\right)^{-1}}$$

168 Define p^{t*} and q^{t*} as the period t price and quantity vectors that include only products that are present in both periods.

169 The matched product expenditure shares defined by (11.195), $s_{tn}^* \equiv v_{tn} / v_t^*$, differ from the original "true" expenditure shares defined as $s_{tn} \equiv v_{tn} / v_t$ because the true period t expenditures v_t include expenditures on "isolated" products that are present in only one of the two periods under consideration. Thus, if there are isolated products in both periods, v^t will be greater than v^{t*} for $t = 1, 2$ and thus the two sets of shares will be different.

$$= \frac{\sum_{n \in S^*} h\left(s_{1n}^*, s_{2n}^*\right) \ln\left(p_{2n} / p_{1n}\right)}{\sum_{i \in S^*} h\left(s_{1i}^*, s_{2i}^*\right)} \tag{11.197}$$

where $h\left(s_{1n}^*, s_{2n}^*\right)$ is the harmonic mean of the restricted shares s_{1n}^* and s_{2n}^*. Thus $P_{\text{WTPDM}}\left(p^{1*}, p^{2*}, q^{1*}, q^{2*}\right) \equiv \exp\left[\rho_2^{**}\right]$ where ρ_2^{**} is defined by (11.197).

The relationship between the *true shares*, the s_{tn}, and the *restricted shares*, the s_{tn}^*, for the always present products is given by the following equations:

$$s_{tn} \equiv \frac{v_{tn}}{v_t} = \frac{v_{tn}}{v_t^*} \frac{v_t^*}{v_t} = s_{tn}^* f_t; \quad t = 1, 2; \quad n \in S^* \tag{11.198}$$

.

where the *fraction* of expenditures on always available commodities compared to expenditures on all commodities during period t is $f_t \equiv v_t^* / v_t$ for $t = 1, 2$. Using definitions (11.190) and (11.198), it can be seen that the logarithm of $P_{\text{WTPD}}\left(p^1, p^2, q^1, q^2\right)$ defined by (11.192) is equal to the following expression:

$$\rho_2^* \equiv \frac{\sum_{n \in S^*} h(s_{1n}, s_{2n}) \ln\left(\frac{p_{2n}}{p_{1n}}\right)}{\sum_{i \in S^*} h(s_{1i}, s_{2i})}$$

$$= \frac{\sum_{n \in S^*} h\left(f_1 s_{1n}^*, f_2 s_{2n}^*\right) \ln\left(\frac{p_{2n}}{p_{1n}}\right)}{\sum_{i \in S^*} h\left(f_1 s_{1i}^*, f_2 s_{2i}^*\right)}. \tag{11.199}$$

Now compare (11.197) and (11.199). If either: (i) $p_{2n} = \lambda p_{1n}$ for all $n \in S^*$ so that we have price proportionality for the always present products or (ii) $f_1 = f_2$ so that the ratio of expenditures on always present products to total expenditure in each period is constant across the two periods, then $\rho_2^{**} = \rho_2^*$. However, if these conditions are not satisfied and there is considerable variation in prices and quantities across periods, then ρ_2^{**} could differ substantially from ρ_2^*. Since neither index is superlative, it is difficult to recommend one of these indexes over the other as the "optimal" carry forward and backward inflation rate that could be used to construct the inflation adjusted carry forward and backward estimates for the missing prices.[170]

In the following section, we define weighted time dummy regression models for the general case of T periods and missing observations.

Weighted Time Product Dummy Regressions: The General Case

We first consider the case of T periods and no missing observations. The generalization of the two-period weighted least squares minimization problem that was defined by (11.175) in section "Weighted Time Product Dummy Regressions: The Bilateral Case" to the case of $T > 2$ periods is (11.200)[171]:

$$\min_{\rho,\beta}\left\{\sum_{n=1}^{N}\sum_{t=1}^{T} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2\right\}. \tag{11.200}$$

The first-order necessary conditions for $\rho^* \equiv [\rho_1^*, \ldots, \rho_T^*]$ and $\beta^* \equiv [\beta_1^*, \ldots, \beta_N^*]$ to solve (11.200) are the following T Eq. (11.201) and N Eq. (11.202):

$$\rho_t^* = \sum_{n=1}^{N} s_{tn}[\ln p_{tn}^* - \beta_n^*]; \quad t = 1, \ldots, T; \tag{11.201}$$

$$\beta_n^* = \frac{\sum_{t=1}^{T} s_{tn}[\ln p_{tn}^* - \rho_t^*]}{\sum_{t=1}^{T} s_{tn}}; \quad n = 1, \ldots, N. \tag{11.202}$$

As usual, the solution to (11.200) given by (11.201) and (11.202) is not unique: if $\rho^* \equiv [\rho_1^*, \ldots, \rho_T^*]$ and $\beta^* \equiv [\beta_1^*, \ldots, \beta_N^*]$ solve (11.201) and

[170] For another alternative weighting scheme for a bilateral time product dummy model in the case of two periods that generalizes the model defined by (11.179) to the case of missing observations, see de Haan (2004a).

[171] Rao (1995, 2004, 2005, 574) was the first to consider this model using expenditure share weights. However, Balk (1980, 70) suggested this class of models much earlier using somewhat different weights.

(11.202), then so do $[\rho_1^* + \lambda, \ldots, \rho_T^* + \lambda]$ and $[\beta_1^* - \lambda, \ldots, \beta_N^* - \lambda]$ for all λ. Thus we can set $\rho_1^* = 0$ in Eq. (11.201) and drop the first equation in (11.201) and use linear algebra to find a unique solution for the resulting equations.[172] Once the solution is found, define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* as follows:

$$\pi_t^* \equiv \exp[\rho_t^*];\ t = 2, 3, \ldots, T \quad \alpha_n^* \equiv \exp[\beta_n^*];\ n = 1, \ldots, N. \quad (11.203)$$

Note that the resulting *price index* between periods t and τ is equal to the following expression:

$$\frac{\pi_t^*}{\pi_\tau^*} = \frac{\prod_{n=1}^{N} \exp\left[s_{tn} \ln\left(\frac{p_{tn}}{\alpha_n^*}\right)\right]}{\prod_{n=1}^{N} \exp\left[s_{\tau n} \ln\left(\frac{p_{\tau n}}{\alpha_n^*}\right)\right]}; \quad 1 \le t, \tau \le T. \quad (11.204)$$

If $s_{tn} = s_{\tau n}$ for $n = 1, \ldots, N$, then π_t^*/π_r^* will equal a weighted geometric mean of the price ratios $p_{tn}/p_{\tau n}$ where the weight for $p_{tn}/p_{\tau n}$ is the common expenditure share $s_{tn} = s_{\tau n}$. Thus π_t^*/π_τ^* will not depend on the α_n^* in this case.[173]

The price levels π_t^* defined by (11.203) are functions of the T price vectors, $p^1, \ldots, p^T$ and the T quantity vectors $q^1, \ldots, q^T$. These price level functions have some good axiomatic properties: (i) the π_t^* are invariant to changes in the units of measurement; (ii) π_t^* regarded as a function of the period t price vector p^t is linearly homogeneous in the components of p^t; i.e., $\pi_t^*(\lambda p^t) = \lambda \pi_t^*(p^t)$ for all $p^t >> 0_N$ and $\lambda > 0$; (iii) π_t^* regarded as a function of the period t quantity vector q^t is homogeneous of degree 0 in the components of q^t; i.e., $\pi_t^*(\lambda q^t) = \pi_t^*(q^t)$ for all $q^t >> 0_N$ and $\lambda > 0$;[174] (iv) the π_t^* satisfy a version of Walsh's (1901,

[172] Alternatively, one can set up the linear regression model defined by $(s_{tn})^{1/2} \ln p_{tn} = (s_{tn})^{1/2}\rho_t + (s_{tn})^{1/2}\beta_n + e_{tn}$ for $t = 1, \ldots, T$ and $n = 1, \ldots, N$ where we set $\rho_1 = 0$ to avoid exact multicollinearity. Iterating between Eqs. (11.201) and (11.202) will also generate a solution to these equations and the solution can be normalized so that $\rho_1 = 0$.

[173] This case is consistent with utility maximizing purchasers having common Cobb Douglas preferences.

[174] By looking at the minimization problem defined by (11.200), it is also straightforward to show that $\pi_t^*(\lambda q^\tau) = \pi_t^*(q^\tau)$ for all $q^\tau >> 0_N$ and $\lambda > 0$ for $\tau = 1, \ldots, T$.

389; 1921b, 540) *multiperiod identity test*; i.e., if $p^t = p^\tau$ and $q^t = q^\tau$, then $\pi_t^* = \pi_\tau^*$.[175]

Once the estimates for the π_t and α_n have been computed, we have the usual two methods for constructing period by period price and quantity levels, P^t and Q^t for $t = 1, \ldots, T$. The π_t^* estimates can be used to form the aggregates using Eq. (11.205) or the α_n^* estimates can be used to form the aggregates using Eq. (11.206)[176]:

$$P^{t*} \equiv \pi_t^*; \quad Q^{t*} \equiv \frac{p^t \cdot q^t}{\pi_t^*}; \quad t = 1, \ldots, T; \tag{11.205}$$

$$Q^{t**} \equiv \alpha^* \cdot q^t; \quad P^{t**} \equiv \frac{p^t \cdot q^t}{\alpha^* \cdot q^t}; \quad t = 1, \ldots, T. \tag{11.206}$$

Define the error terms $e_{tn} \equiv \ln p_{tn} - \ln \pi_t^* - \ln \alpha_n^*$ for $t = 1, \ldots, T$ and $n = 1, \ldots, N$. If all $e_{tn} = 0$, then P^{t*} will equal P^{t**} and Q^{t*} will equal Q^{t**} for $t = 1, \ldots, T$. However, if the error terms are not all equal to zero, then the statistical agency will have to decide on pragmatic grounds on which option to choose.

It is straightforward to generalize the weighted least squares minimization problem (11.200) to the case where there are missing prices and quantities. As in section "Time Product Dummy Regressions: The Case of Missing Observations" we assume that there are N products and T time periods but not all products are purchased (or sold) in all time periods. For each period t, define the set of products n that are present in period t as $S(t) \equiv \{n : p_{tn} > 0\}$ for $t = 1, 2, \ldots, T$. It is assumed that these sets are not empty; i.e., at least one product is purchased in each period. For each product n, define the set of periods t where product n is present as $S^*(n) \equiv \{t : p_{tn} > 0\}$. Again, assume that these sets are not empty; i.e., each product is sold in at least one time period. The generalization of (11.200) to the case of missing products is the following weighted least

[175] We would like the π_t^* to satisfy the usual (strong) identity test, which is: if $p^t = p^\tau$, then $\pi_t^* = \pi_\tau^*$. However, if the share weights for the two periods are different, then this test no longer holds. However, if we define the period t price and quantity levels using definitions (11.206), it can be seen that the resulting Q^{t**} will satisfy the usual (strong) identity test for quantities. If our perspective is one of measuring economic welfare, then we may want to choose (11.206) over (11.205).

[176] Note that the price level P^{t**} defined in (11.206) is a quality adjusted unit value index of the type studied by de Haan (2004b).

squares minimization problem:

$$\min_{\rho,\beta} \sum_{t=1}^{T} \sum_{n \in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2 = \min_{\rho,\beta} \sum_{n=1}^{N} \sum_{t \in S^*(n)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2. \tag{11.207}$$

.

Note that there are two equivalent ways of writing the least squares minimization problem. The first-order necessary conditions for $\rho_1, \ldots, \rho_T$ and $\beta_1, \ldots, \beta_N$ to solve (11.207) are the following counterparts to (11.201) and (11.202)[177]:

$$\sum_{n \in S(t)} s_{tn}\left[\rho_t^* + \beta_n^*\right] = \sum_{n \in S(t)} s_{tn} \ln p_{tn}; \quad t = 1, \ldots, T; \tag{11.208}$$

$$\sum_{t \in S^*(n)} s_{tn}\left[\rho_t^* + \beta_n^*\right] = \sum_{t \in S^*(n)} s_{tn} \ln p_{tn}; n = 1, \ldots, N. \tag{11.209}$$

As usual, the solution to (11.208) and (11.209) is not unique: if $\rho^* \equiv \left[\rho_1^*, \ldots, \rho_T^*\right]$ and $\beta^* \equiv \left[\beta_1^*, \ldots, \beta_N^*\right]$ solve (11.208) and (11.209), then so do $\left[\rho_1^* + \lambda, \ldots, \rho_T^* + \lambda\right]$ and $\left[\beta_1^* - \lambda, \ldots, \beta_N^* - \lambda\right]$ for all λ. Thus we can set $\rho_1^* = 0$ in Eq. (11.208) and drop the first equation in (11.208) and use linear algebra to find a unique solution for the resulting equations.

Define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* by definitions (11.203). The Weighted Time Product Dummy price level for period t is defined as $p_{\text{WTPD}}^t \equiv \pi_t^*$ for $t = 1, \ldots, T$. Substitute these definitions into Eqs. (11.208) and (11.209). After some rearrangement, Eqs. (11.208) and (11.209) become the following ones:

$$\pi_t^* = \exp\left[\sum_{n \in S(t)} s_{tn} \ln\left(\frac{p_{tn}}{\alpha_n^*}\right)\right] \equiv p_{\text{WTPD}}^t; \quad t = 1, \ldots, T; \tag{11.210}$$

[177] Equations (11.208) and (11.209) show that the solution to (11.207) does not depend on any independently determined reservation prices p_{tn} for products n that are missing in period t.

$$\alpha_n^* = \exp\left[\frac{\sum_{t \in S^*(n)} s_{tn} \ln\left(\frac{p_{tn}}{\pi_t^*}\right)}{\sum_{t \in S^*(n)} s_{tn}}\right]; \quad n = 1, \ldots, N. \tag{11.211}$$

Once the estimates for the π_t and α_n have been computed, we have the usual two methods for constructing period by period price and quantity levels, P^t and Q^t for $t = 1, \ldots, T$; see (11.205) and (11.206) above.[178]

The new price levels π_t^* defined by (11.210) are functions of the T price vectors, $p^1, \ldots, p^T$ and the T quantity vectors $q^1, \ldots, q^T$. If there are missing products, the corresponding prices and quantities, p_{tn} and q_{tn}, are temporarily set equal to 0. The new price level functions defined by (11.210) have the same axiomatic properties (i)–(iv) which were noted earlier in this section.[179] The present price level functions take the economic importance of the products into account and thus are a clear improvement over their unweighted counterparts which were discussed in

[178] The counterparts to definitions (11.205) are now: $P^{t*} \equiv \pi_t^* = \prod_{n \in S(t)} \exp[s_{tn} \ln(p_{tn}/\alpha_n^*)]$, a share weighted geometric mean of the quality adjusted prices present in period t, and $Q^{t*} \equiv \sum_{n \in S(t)} (p_{tn}q_{tn}/P^{t*})$ for $t = 1, \ldots, T$. The counterparts to Eq. (11.206) are now: $Q^{t**} \equiv \sum_{n \in S(t)} \alpha_n^* q_{tn}$ and $P^{t**} \equiv \sum_{n \in S(t)} (p_{tn}q_{tn}/Q^{t**}) = \sum_{n \in S(t)} p_{tn}q_{tn} \Big/ \sum_{n \in S(t)} \alpha_n^* q_{tn}$ $= \sum_{n \in S(t)} p_{tn}q_{tn} \Big/ \sum_{n \in S(t)} \alpha_n^* \left(p_{tn}^{-1}\right) p_{tn}q_{tn} = \left[\sum_{n \in S(t)} s_{tn}\left(p_{tn}/\alpha_t^*\right)^{-1}\right]^{-1}$, a share weighted harmonic mean of the quality adjusted prices present in period t. Thus using Schlömilch's inequality (see Hardy et al. [1934, 26]), we see that $P^{t**} \leq P^{t*}$ which in turn implies that $Q^{t**} \geq Q^{t*}$ for $t = 1, \ldots, T$. This algebra is due to de Haan (2004b) (2010) and de Haan and Krsinich (2018, 763). If the variance of prices increases over time, it is likely that P^{t**}/P^{t*} will be less that $P^{t*}\Big/P^{1*}$ and vice versa if the variance of prices decreases; see de Haan and Krsinich (2018, 771) and Diewert (2018, 10) on this last point. Note that the work of de Haan and Krsinich provides us with a concrete formula for the difference between P^{t*} and P^{t**}. The model used by de Haan and Krsinich is a more general hedonic regression model which includes the time dummy model used in the present section as a special case.

[179] However, we would like the P^{t*} to satisfy a strong identity test as noted above; i.e., we would like P^{t*} to equal P^{r*} if the prices in periods t and r are identical. The P^{t*} equal to the π_t^* where the π_t^* are defined by (11.210) do not satisfy this strong identity test for price levels. However, the Q^{t**} defined as $\sum_{n \in S(t)} \alpha_n^* q_{tn}$ do satisfy the strong identity test for quantities and this suggests that the P^{t**}, Q^{t**} decomposition of period t sales may be a better choice than the P^{t*}, Q^{t*} decomposition.

section "Time Product Dummy Regressions: The Case of Missing Observations". If the estimated errors $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ that implicitly appear in the weighted least squares minimization problem (11.207) turn out to be small, then the underlying exact model, $p_{tn} = \pi_t \alpha_n$ for $t = 1, \ldots, T$, $n \in S(t)$, provides a good approximation to reality and thus this weighted time product dummy regression model can be used with some confidence.

The solution to the weighted least squares minimization problem defined by (11.207), π_t^* for $t = 1, \ldots, T$ and α_n^* for $n = 1, \ldots, N$ can be used to define (neutral) *reservation prices* for missing observations. For any missing price for product n in period t, define p_{tn}^* as follows:

$$p_{tn}^* \equiv \pi_t^* \alpha_n^*; \quad n \in S^*(t). \tag{11.212}$$

In what follows, we will use the prices defined by (11.212) to replace the 0 prices in the vectors p^t for $t = 1, \ldots, T$ so with the use of these imputed prices, all price vectors p^t have positive components. Of course, the quantities q_{tn} and the shares s_{tn} that correspond to the imputed prices defined by (11.212) are still equal to 0.

The weighted time product dummy price level functions p_{WTPD}^t defined by (11.210) have the same unsatisfactory property that their unweighted counterparts had in previous sections: a product that is available only in one period out of the T periods has no influence on the aggregate price levels $p_{\text{WTPD}}^t \equiv \pi_t^*$.[180] This means that the price of a new product that appears in period T has no influence on the price levels and thus the benefits of an expanding consumption set are not measured by this multilateral method. This is a significant shortcoming of this method. However, on the positive side of the ledger, this method does satisfy the strong identity test for the companion quantity index, a property that it shares with the GK multilateral method.[181]

[180] See Diewert (2004) for a proof or modify the proof in section "Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations".

[181] Both methods are basically quality adjusted unit value methods. Thus, if the products under consideration are highly substitutable, then both methods may give satisfactory results. From the viewpoint of the economic approach to index number theory, the GK method is consistent with utility maximizing behavior if purchasers have either Leontief (no substitution) preferences or linear preferences (perfect substitution preferences after quality adjustment). The weighted time product dummy method is consistent with utility

Once the WTPD price levels p^t_{WTPD} have been defined,[182] the *weighted time product dummy price index* for period t (relative to period 1) is defined as $P^t_{\text{WTPD}} \equiv p^t_{\text{WTPD}} / p^1_{\text{WTPD}}$ and the logarithm of P^t_{WTPD} is equal to the following expression:

$$\ln P^t_{\text{WTPD}} = \sum_{n=1}^{N} s_{tn}\left(\ln p_{tn} - \beta_n^*\right) - \sum_{n=1}^{N} s_{1n}\left(\ln p_{1n} - \beta_n^*\right); \quad t = 1, \ldots, T. \tag{11.213}$$

.
With the above expression for $\ln P^t_{\text{WTPD}}$ in hand, we can compare $\ln P^t_{\text{WTPD}}$ to $\ln P^t_T$. Using (11.213) and definition (11.40),[183] we can derive the following expressions for $t = 1, \ldots, T$:

$$\ln P^t_{\text{WTPD}} - \ln P^t_T = \left(\frac{1}{2}\right) \sum_{n=1}^{N} (s_{tn} - s_{1n})\left(\ln p_{tn} - \beta_n^*\right) + \left(\frac{1}{2}\right) \sum_{n=1}^{N} (s_{tn} - s_{1n})\left(\ln p_{1n} - \beta_n^*\right). \tag{11.214}$$

Since $\sum_{n=1}^{N} (s_{tn} - s_{1n}) = 0$ for each t, the two sets of terms on the right hand side of equation t in (11.214) can be interpreted as normalizations of the covariances between the vectors $s^t - s^1$ and $\ln p^t - \beta^*$ for the first set of terms and between $s^t - s^1$ and $\ln p^1 - \beta^*$ for the second set of terms. If the products are highly substitutable with each other, then a low p_{tn} will usually imply that $\ln p_{tn}$ is less than the average log price for product n, β_n^*, and it is also likely that s_{tn} is greater than s_{1n} so that $(s_{tn} - s_{1n})\left(\ln p_{tn} - \beta_n^*\right)$ is likely to be negative. Hence, the covariance

maximizing behavior if purchasers have either Cobb Douglas preferences or linear preferences. Note that Cobb Douglas preferences are not consistent with situations where there are new and disappearing products.

[182] See Eq. (11.210).

[183] If product n in period t is missing, we use the imputed price p^*_{tn} defined by (11.212) as the positive reservation price for this observation in the definitions for both P^t_{WTPD} and P^t_T which appear in Eqs. (11.213) and (11.214). Thus, the summations in (11.213) and (11.214) are over all N products.

between $s^t - s^1$ and $\ln p^t - \beta^*$ will tend to be negative. On the other hand, if p_{1n} is unusually low, then $\ln p_{1n}$ will be less than the average log price β_n^* and it is likely that s_{1n} is greater than s_{tn} so that $(s_{tn} - s_{1n})(\ln p_{1n} - \beta_n^*)$ is likely to be positive. Hence, the covariance between $s^t - s^1$ and $\ln p^1 - \beta^*$ will tend to be positive. Thus, the first set of terms on the right hand side of (11.214) will tend to be negative while the second set will tend to be positive. If there are no divergent trends in log prices and sales shares, then it is likely that these two terms will largely offset each other and under these conditions, P_{WTPD}^t is likely to approximate P_T^t reasonably well. However, with divergent trends and highly substitutable products, it is likely that the first set of negative terms will be larger in magnitude than the second set of terms and thus P_{WTPD}^t is likely to be below P_T^t under these conditions.[184] But if some product n is not available in period 1 so that $s_{1n} = 0$ and if the logarithm of the imputed price for this product p_{1n}^* defined by (11.212) is greater than β_n^*, then it can happen that the second covariance term on the right hand side of (11.214) becomes very large and positive so that it overwhelms the first negative covariance term and thus P_{WTPD}^t ends up above P_T^t rather than below it.

To sum up, the weighted time product indexes can be problematic in the elementary index context when price and quantity data are available as compared to a fixed base superlative index (that uses reservation prices):

- If there are no missing products and the products are strong substitutes, the WTPD indexes will tend to have a downward bias.
- If there are no missing products and the products are weak substitutes, the WTPD indexes will tend to have an upward bias.
- If there are missing products in period 1, the relationship between the WTPD indexes and the corresponding Törnqvist Theil indexes is uncertain.
- If there are missing products, the weighted time product dummy price levels and price indexes do not depend on reservation prices

[184] If the products are not highly substitutable so that when a price goes up, the quantity purchased goes down but the expenditure share also goes up, then the inequalities are reversed; i.e., if there are no missing products and long-term trends in prices and quantities, then P_{WTPD}^t is likely to be above P_T^t. If preferences of purchasers are Cobb Douglas, then expenditure shares will remain constant over time and P_{WTPD}^t will equal P_T^t for $t = 1, \ldots, T$.

(which could be regarded as an advantage of the WTPD indexes for price statisticians who want to avoid making imputations).

Linking Based on Relative Price Similarity

The GEKS multilateral method treats each set of price indexes using the prices of one period as the base period as being equally valid and hence an averaging of the resulting parities seems to be appropriate under this hypothesis. Thus, the method is "democratic" in that each bilateral index number comparison between any two periods gets the same weight in the overall method. However, it is not the case that all bilateral comparisons of price between two periods are equally accurate: if the relative prices in periods r and t are very similar, then the Laspeyres and Paasche price indexes will be very close to each other and hence it is likely that the "true" price comparison between these two periods (using the economic approach to index number theory) will be very close to the bilateral Fisher index that compares prices between the two periods under consideration. In particular, if the two price vectors are exactly proportional, then we want the price index between these two periods to be equal to the factor of proportionality and the direct Fisher index between these two periods satisfies this proportionality test. On the other hand, the GEKS index comparison between the two periods would not in general satisfy this proportionality test.[185] Also if prices are identical between two periods but the quantity vectors are different, then GEKS price index between the two periods would not equal unity in general.[186] The above considerations suggest that a more accurate set of price indexes could be constructed if initially a bilateral comparison was made between the two periods that have the most *similar relative price structures*. At the next stage of the comparison, look for a third period that had the most similar relative price structure to the first two periods and link in this third

[185] If both prices *and* quantities are proportional to each other for the two periods being compared, then the GEKS price index between the two periods will satisfy this (weak) proportionality test. However, we would like the GEKS price index between the two periods to satisfy the strong proportionality test; i.e., if the two price vectors are proportional (and the two quantity vectors are not necessarily proportional to each other), then we would like the GEKS price index between the two periods to equal the factor of proportionality.

[186] See Zhang et al. (2019, 689) on this point.

country to the comparisons of volume between the first two countries and so on. At the end of this procedure, a pathway through the periods in the window would be constructed, that minimized the sum of the relative price dissimilarity measures. In the context of making comparisons of prices across countries, this method of linking countries with the most similar structure of relative prices has been pursued by Hill (1997, 1999a, 1999b, 2009), Hill and Timmer (2006), Diewert (2009, 2013, 2018) and Hill et al. (2017). Hill (2001, 2004) also pursued this similarity of relative prices approach in the time series context. Our conclusion is that similarity linking using Fisher ideal price indexes as the bilateral links is an attractive alternative to GEKS.

A key aspect of this methodology is the choice of the measure of similarity (or dissimilarity) of the relative price structures of two countries. Various measures of the similarity or dissimilarity of relative price structures have been proposed by Allen and Diewert (1981), Kravis et al. (1982, 104–106), Hill (1997, 2009), Sergeev (2001, 2009), Hill and Timmer (2006), Aten and Heston (2009), and Diewert (2009, 2021a).

In this section, we will discuss the following *weighted asymptotic linear index of relative price dissimilarity*, Δ_{AL}, suggested by Diewert (2009)[187]:

$$\Delta_{AL}(p^r, p^t, q^r, q^t) \equiv \sum_{n=1}^{N}\left(\frac{1}{2}\right)(s_{rn}+s_{tn})\left\{\left(\frac{p_{tn}}{P_F(p^r, p^t, q^r, q^t)p_{rn}}\right)+\left(\frac{P_F(p^r, p^t, q^r, q^t)p_{rn}}{p_{tn}}\right)-2\right\} \tag{11.215}$$

where $P_F(p^r, p^t, q^r, q^t) \equiv \left[p^t \cdot q^r p^t \cdot q^t / p^r \cdot q^r p^r \cdot q^t\right]^{1/2}$ is the bilateral Fisher price index linking period t to period r and p^r, q^r, s^r and p^t, q^t, s^t are the price, quantity and share vectors for periods r and t respectively. This measure turns out to be nonnegative and the bigger $\Delta_{AL}(p^r, p^t, q^r, q^t)$ is, the more dissimilar are the relative prices for periods r and t. Note that if $p^t = \lambda p^r$ for some positive scalar so that if prices are proportional for the two periods, then $\Delta_{AL}(p^r, p^t, q^r, q^t) = 0$. Note also that *all prices need to be positive* in order for $\Delta_{AL}(p^r, p^t, q^r, q^t)$ to be well defined. Thus, if there are missing products in one of the two periods being compared, reservation prices need to be estimated for the

[187] The discussion paper version of Diewert (2009) appeared in (2002).

missing product prices in each period.[188] Alternatively, inflation adjusted carry forward or carry backward prices can be used to fill in the missing prices.[189]

The method for constructing *Similarity Linked Fisher* price indexes in real time using the above measure of relative price similarity proceeds as follows. Set the similarity linked price index for period 1, $P_{AL}^1 \equiv 1$. The period 2 index is set equal to $P_F(p^1, p^2, q^1, q^2)$, the Fisher index linking the period 2 prices to the period 1 prices. Thus $P_{AL}^2 \equiv P_F(p^1, p^2, q^1, q^2)P_{AL}^1$. For period 3, evaluate the dissimilarity indexes $\Delta_{AL}(p^1, p^3, q^1, q^3)$ and $\Delta_{AL}(p^2, p^3, q^2, q^3)$ defined by (11.215). If $\Delta_{AL}(p^1, p^3, q^1, q^3)$ is the minimum of the two numbers, $\Delta_{AL}(p^1, p^3, q^1, q^3)$ and $\Delta_{AL}(p^2, p^3, q^2, q^3)$, define $P_{AL}^3 \equiv P_F(p^1, p^3, q^1, q^3)P_{AL}^1$. If $\Delta_{AL}(p^2, p^3, q^2, q^3)$ is the minimum of these two numbers, define $P_{AL}^3 \equiv P_F(p^2, p^3, q^2, q^3)P_{AL}^2$. For period 4, evaluate the dissimilarity indexes $\Delta_{AL}(p^r, p^4, q^r, q^4)$ for $r = 1, 2, 3$. Let r^* be such that $\Delta_{AL}(p^{r*}, p^4, q^{r*}, q^4) = \min_r\{\Delta_{AL}(p^r, p^4, q^r, q^4); r = 1, 2, 3\}$.[190] Then define $P_{AL}^4 \equiv P_F(p^{r*}, p^4, q^{r*}, q^4)P_{AL}^{r*}$. Continue this process in the same manner; i.e., for period t, let r^* be such that $\Delta_{AL}(p^{r*}, p^t, q^{r*}, q^t) = \min_r\{\Delta_{AL}(p^r, p^t, q^r, q^t); r = 1, 2, \ldots, t-1\}$ and define $P_{AL}^t \equiv P_F(p^{r*}, p^t, q^{r*}, q^t)P_{AL}^{r*}$. This procedure allows for the construction of similarity linked indexes in real time.

Diewert (2018) implemented the above procedure with a retail outlet scanner data set and compared the resulting similarity linked index, P_{AL}^t, to other indexes that are based on the use of superlative indexes and the economic approach to index number theory. The data set he used is listed in section A1 of the Appendix and his results are listed in the Appendix along with some additional results. The comparison indexes in his study were the fixed base Fisher and Törnqvist indexes, P_F^t and P_T^t, and

[188] See section "Time Product Dummy Regressions: The Case of Missing Observations" of Diewert (2021a) for additional information on reservation prices.

[189] See the discussion in the following section. Section A6 of the Appendix compares P_{AL}^t computed using reservation prices and P_{ALC}^t which uses inflation adjusted carry forward/backward prices for missing products. For our particular empirical example, there were small differences in the resulting indexes.

[190] If the minimum occurs at more than one r, choose r^* to be the earliest of these minimizing periods.

the multilateral indexes, P^t_{GEKS} and P^t_{CCDI}. The sample means for these five indexes, P^t_{AL}, P^t_F, P^t_T, P^t_{GEKS} and P^t_{CCDI}, were 0.97069, 0.97434, 0.97607, 0.97417 and 0.97602. Thus on average, P^t_{AL} was about 0.5 percentage points below P^t_T and P^t_{CCDI} and about 0.35 percentage points below P^t_F and P^t_{GEKS}. These are fairly significant differences.[191]

What are some of the advantages and disadvantages of using either P^t_{AL}, P^t_F, P^t_T, P^t_{GEKS} or P^t_{CCDI} as target indexes for an elementary index in a CPI? All of these indexes are equally consistent with the economic approach to index number theory. The problem with the fixed base Fisher and Törnqvist indexes is that they depend too heavily on the base period. Moreover, sample attrition means that the base must be changed fairly frequently, leading to a potential chain drift problem. The GEKS and CCDI indexes also suffer from problems associated with the existence of seasonal products: it makes little sense to include bilateral indexes between all possible periods in a window of periods in the context of seasonal commodities. The similarity linked indexes address both the problem of sample attrition and the problem of seasonal commodities. Moreover, Walsh's multiperiod identity test is always satisfied using this methodology. Finally, there is no need to choose a window length and use a rolling window approach to construct the time series of indexes if the price similarity linking method is used: the window length simply grows by one period as the data for an additional period becomes available.[192]

The procedure for constructing the time series of similarity linked Fisher price indexes, P^t_{AL}, is a *real-time procedure*; i.e., there is no preliminary time period that is required in order to produce the final time series of aggregate price levels. However, the resulting pattern of bilateral links may not be "optimal" in the sense that the most similar sets of relative prices are linked to one another in the first year or so. This is apparent when the price level P^2_{AL} is constructed: it is simply equal to the Fisher index linking period 2 to 1; there are no other choices for a linking

[191] The final values for the five indexes $\left(P^t_{\text{AL}}, P^t_F, P^t_T, P^t_{\text{GEKS}} \text{ and } P^t_{\text{CCDI}}\right)$ were as follows: 0.92575, 0.95071, 0.95482, 0.94591 and 0.94834. Thus P^t_{AL} ended up significantly below the other indexes. P^t_T is listed in Table A.4 and the remaining indexes are listed in Table A.6 of the Appendix.

[192] In practice, as the number of periods grow and the structure of the economy evolves, it will become increasingly unlikely that a current observation will be linked to a distant observation. Thus eventually, it becomes practical to move to a rolling window framework with a large window length.

partner. A "better" set of bilateral links could potentially be obtained if a final set of bilateral links for the index could be obtained by forming a *spanning tree of comparisons* say for the first year of data.[193] Thus a year of data on prices and quantities is used to form a set of bilateral links that minimizes the sum of the associated dissimilarity measures that link the observations for the first year. This leads to a *modified* set of price levels for the first year, say P_{ALM}^t for t in the first year. For months t that follow after the first "training" year, the bilateral links are the same as indicated earlier but because the levels in the first year may have changed, the modified price levels P_{ALM}^t for months t that follow after the first year may differ from the real-time price levels P_{AL}^t described earlier. However, the trends in the two series will be similar. In section A5 of the Appendix, we calculate both P_{AL}^t and P_{ALM}^t for the data set listed in section A1 of the Appendix. There is little difference in these two series for our example data set and in fact, both series end up at the same point.[194] Normally, we do not expect much difference between the original real-time method and the modified method but the modified method is useful in the context of constructing price indexes for strongly seasonal commodities because it will tend to reduce the magnitude of seasonal fluctuations.

Similarity linked price indexes suffer from at least two problems:

- A measure of relative price dissimilarity must be chosen and there may be many "reasonable" choices for the measure of dissimilarity. These different choices can lead to different indexes, which in turn can lead users to question the usefulness of the method.
- The measures of weighted price dissimilarity suggested by Diewert (2009) require that all prices in the comparison of prices between two periods be positive.

[193] See Hill (2001, 2004) for explanations of how this can be done.

[194] See Table A.7 and Chart 9 in the Appendix. Although P_{AL}^t and P_{ALM}^t end up at the same level, the mean of the P_{AL}^t was 0.97069 and the mean of the P_{ALM}^t was 0.96437. The fluctuations in the P_{ALM}^t series were somewhat smaller. This tendency for the modified series to be a bit smoother than the corresponding real-time series becomes important in the context of constructing indexes for strongly seasonal commodities. In this context, the use of the modified similarity linking method is recommended in order to reduce seasonal fluctuations.

These problems will be addressed in section "Linking Based on Relative Price and Quantity Similarity" where an alternative measure of price dissimilarity that does not require strictly positive prices will be defined. Using the scanner data set listed in section A1 of the Appendix, this new measure of price (and quantity) dissimilarity generates indexes P_{SP}^t that are very similar to the P_{AL}^t indexes discussed in the present section.

It is a difficult econometric exercise to estimate reservation prices and so a simpler method may be required in order to construct imputed prices for missing products in a scanner data set. In the following section, a standard method used by price statisticians is explained.

Inflation Adjusted Carry Forward and Backward Imputed Prices

When constructing elementary indexes, statistical agencies often encounter situations where a product in an elementary index disappears. At the time of disappearance, it is unknown whether the product is temporarily unavailable so the missing price could be set equal to the last available price; i.e., the missing price could be replaced by a carry *forward price*. Thus, carry forward prices could be used in place of reservation prices, which are much more difficult to construct. This procedure is, in general, not a recommended one. A much better alternative to the use of a carry forward price is an *inflation adjusted carry forward price*; i.e., the last available price is escalated using the maximum overlap index between the period when the product was last available and the current period where an appropriate index number formula is used.[195] In this section, we use inflation adjusted carry forward and carry backward prices in place of the reservation prices for our scanner data set and compare the resulting indexes with our earlier indexes that used the econometrically estimated reservation prices that were constructed by Diewert and Feenstra (2017) for the scanner data set listed in Appendix 1.

Suppose we have price and quantity data for N products for T periods as usual. Let $p^t \equiv [p_{1t}, \ldots, p_{tN}]$ and $q^t \equiv [q_{t1}, \ldots, q_{tN}]$ denote the

[195] Triplett (2004, 21–29) calls these two methods for replacing missing prices the *link to show no change method* and the *deletion method*. See section "Time Product Dummy Regressions: The Case of Missing Observations" in Diewert (2021a) and Diewert et al. (2017) for a more extensive discussion on the problems associated with finding replacements for missing prices.

period t price and quantity vectors. If product n is not present in period t, define (for now) the corresponding p_{tn} and q_{tn} to be 0. Define $S(t)$ to be the set of products that are present in period t, i.e., $S(t) \equiv \{n : p_{tn} > 0\}$.[196] Suppose that we want to make a Fisher index number comparison between periods r and t where $r < t$. The *maximum overlap set of products* that are present in periods r and t is the intersection set, $S(r) \cap S(t)$. We assume that this set is nonempty. Define the vectors p^{r*}, p^{t*}, q^{r*}, q^{t*} as the vectors that have only the products that are present in periods r and t. Define the *maximum overlap Fisher price index* for period t relative to period r as $P_{\text{FM}}\left(p^{r*}, p^{t*}, q^{r*}, q^{t*}\right)$. If there are products present in period r that are not present in period t, define the *inflation adjusted carry forward price* for such products as follows:

$$p_{tn} \equiv p_{rn} P_{\text{FM}}\left(p^{r*}, p^{t*}, q^{r*}, q^{t*}\right); \quad n \in S(r);\ n \notin S(t). \tag{11.216}$$

The corresponding quantities q_{tn} remains at their initially defined 0 levels. If there are products present in period t that are not present in period r, define the *inflation adjusted carry backward price* for such products as follows:

$$p_{rn} \equiv \frac{p_{tn}}{P_{\text{FM}}(p^{r*}, p^{t*}, q^{r*}, q^{t*})}; \quad n \in S(r);\ n \notin S(t). \tag{11.217}$$

The corresponding quantities q_{rn} remain at their initial 0 levels.

Using the above definitions, we will have new price and quantity vectors that have well-defined price and quantity vectors p^{r**}, p^{t**}, q^{r**}, q^{t**} that have positive prices for products that belong to the union set of products that are present in both periods r and t, $S(r) \cup S(t)$. Denote the Fisher index for period t relative to period r over this union set of products as $P_F^*\left(p^{r**}, p^{t**}, q^{r**}, q^{t**}\right)$. This index can be used as the Fisher index linking periods r and t. Thus the carry forward and carry backward prices defined by (11.216) and (11.217) can replace econometrically estimated reservation prices and the similarity linked price indexes defined in the previous section can be calculated using the Fisher linking indexes $P_F^*\left(p^{r**}, p^{t**}, q^{r**}, q^{t**}\right)$ in place of the $P_F\left(p^r, p^t, q^r, q^t\right)$ used in the previous section. Note that the components of the period t price vector p^{t**} will be equal to the components of the original period t price vector p^t except for components that correspond to missing products.

[196] Recall that this notation was used in previous sections.

It should be emphasized that, usually, it is important to make the index number adjustments to the carry forward and backward prices defined by (11.216) and (11.217) instead of simply carrying existing prices from one period to another period. Failure to make these index number adjustments could lead to substantial biases if substantial general inflation (or deflation) is present. From the perspective of the economic approach to index number theory, it is likely that the use of inflation adjusted carry backward prices in place of estimated reservation prices will in general lead to an upward bias in the linking index since the "true" reservation prices are likely to be higher than the adjusted prices in order to induce consumers to purchase zero units of the unavailable products in the prior period. Obviously, the bias in using carry forward prices for disappearing products works in the opposite direction.

In section A6 of the Appendix, we used our scanner data to compute the GEKS, Fisher, Chained Fisher and the real-time similarity linked index explained in the previous section which used the Δ_{AL} dissimilarity measure defined by (11.215). We also calculated the real-time Predicted Share similarity linked indexes that use the Δ_{SP} dissimilarity measure that will be defined by (11.218) in the following section. Denote the resulting period t index by P_{SP}^t. There were missing products in our scanner data set. As noted above, the missing prices were initially set equal to reservation prices calculated using econometrics. Denote these indexes for period t (which used reservation prices) by P_{GEKS}^t, P_F^t, P_{FCH}^t, P_{AL}^t and P_{SP}^t. The same five indexes were recomputed using inflation adjusted carry forward and carry backward prices for the missing product prices.[197] Denote the resulting period t indexes by P_{GEKSC}^t, P_{FC}^t, P_{FCHC}^t, P_{ALC}^t and P_{SPC}^t. As noted earlier, it turns out that Geary Khamis index (P_{GK}^t) and Weighted Time Product Dummy index $\left(P_{WTPD}^t\right)$ do not depend on the values of the missing prices and so these indexes do not have to be recomputed using carry forward prices in place of reservation prices. P_{GK}^t and P_{WTPD}^t are listed in Table A.6 in section A5 of the Appendix. The series P_{AL}^t, P_{ALC}^t, P_{SP}^t, P_{SPC}^t, P_{GEKS}^t, P_{GEKSC}^t are listed in Table A.8 in section A6 of the Appendix A6 along with the Fisher and Chained Fisher indexes using reservation prices, denoted by P_F^t and P_{FCH}^t, and using carry forward prices, denoted by P_{FC}^t and P_{FCHC}^t.

[197] Inflation adjusted carry forward prices were used to compute prices for missing products except when a product was missing in period 1. In the latter case, inflation adjusted carry backward prices were computed for the missing products.

A summary of the results using econometrically estimated reservation prices versus using carry forward and backward prices for the missing products is as follows: for our example, there was very little difference between the resulting index pairs using reservation prices versus using inflation adjusted carry forward prices. This is likely due to the fact that only 20 out of 741 prices were missing; i.e., only 2.7% of the sample had missing products. (0.97542) and $P^{A}_{\mathrm{FCH}} = 1.0589$ (1.0589). Our tentative conclusion here is that *for products that are highly substitutable, the use of inflation adjusted forward and backward prices for missing products will probably generate weighted indexes that are comparable to their counterparts that use econometrically estimated reservation prices*. For products which are not highly substitutable, it is likely that reservation prices will be higher than their inflation adjusted carry forward prices and thus it is likely that the indexes will differ in a more substantial manner. This conclusion is only tentative and further research on the use of reservation prices is required.

Linking Based on Relative Price and Quantity Similarity

A problem with the measure of relative price dissimilarity $\Delta_{\mathrm{AL}}(p^r, p^t, q^r, q^t)$ defined by (11.215) is that it requires that all prices in the two periods being compared must be positive. Thus, if there are missing prices for some products present in one of the two periods but not in the other period, then the Δ_{AL} dissimilarity measure is not well defined.[198]

[198] Diewert (2009, 205–206) recommended two other measures of price dissimilarity but they also have the problem that they are also not well defined if some product prices are equal to 0. These alternative measures are the *weighted log quadratic measure of relative price dissimilarity*, $\Delta_{\mathrm{PLQ}}\left(p^1, p^2, q^1, q^2\right) \equiv \sum_{n=1}^{N}(1/2)\left(s_n^1 + s_n^2\right)\left[\ln\left(p_n^2 / p_n^1 P\left(p^1, p^2, q^1, q^2\right)\right)\right]^2$, and the *weighted asymptotically quadratic measure of relative price dissimilarity*, $\sum_{n=1}^{N}(1/2)\left(s_n^1 + s_n^2\right)\left\{\left[\left(p_n^2 / p_n^1 P\left(p^1, p^2, q^1, q^2\right)\right) - 1\right]^2 + \left[\left(P\left(p^1, p^2, q^1, q^2\right)p_n^1 / p_n^2\right) - 1\right]^2\right\} \equiv \Delta_{\mathrm{WAQ}}\left(p^1, p^2, q^1, q^2\right)$, where $\equiv P\left(p^1, p^2, q^1, q^2\right)$ is any superlative bilateral price index formula. It can be shown that $\Delta_{\mathrm{PLQ}}\left(p^1, p^2, q^1, q^2\right)$ approximates $\Delta_{\mathrm{AL}}(p^r, p^t, q^r, q^t)$ to the second order around any point where $p^1 = p^2 >> 0_N$ and $q^1 = q^2 >> 0_N$.

The following *Predicted Share measure of relative price dissimilarity*, $\Delta_{SP}(p^r, p^t, q^r, q^t)$, is well defined even if some product prices in the two periods being compared are equal to zero:

$$\Delta_{SP}(p^r, p^t, q^r, q^t) \equiv \sum_{n=1}^{N} \left[s_{tn} - \frac{p_{rn}q_{tn}}{p^r \cdot q^t} \right]^2 + \sum_{n=1}^{N} \left[s_{rn} - \frac{p_{tn}q_{rn}}{p^t \cdot q^r} \right]^2 \qquad (11.218)$$

where $s_{tn} \equiv p_{tn}q_{tn}/p^t \cdot q^t$ is the share of product n in period t expenditures on the N products for $t = 1, \ldots, T$ and $n = 1, \ldots, N$. We require that $p^r q^t > 0$ for $r = 1, \ldots, T$ and $t = 1, \ldots, T$ in order for $\Delta_{SP}(p^r, p^t, q^r, q^t)$ to be well defined for any pair of periods, r and t. Since the two summations on the right hand side of (11.218) are sums of squared terms, we see that $\Delta_{SP}(p^r, p^t, q^r, q^t) \geq 0$.

The first set of N terms on the right hand side of (11.218) is $\sum_{n=1}^{N} [s_{tn} - (p_{rn}q_{tn}/p^r \cdot q^t)]^2$. Note that the terms $p_{rn}q_{tn}/p^r \cdot q^t$ for $n = 1, \ldots, N$ are (hybrid) *shares*; i.e., these terms are nonnegative and they sum to unity so that $\sum_{n=1}^{N} (p_{rn}q_{tn}/p^r \cdot q^t) = 1$. These shares use the prices of period r and the quantities of period t. They can be regarded as *predictions* for the actual period t shares, s_{tn}, using the prices of period r but using the quantities of period t. A similar interpretation applies to the second set of N terms on the right hand side of (11.218); the hybrid shares that use the prices of period t and the quantities of period r, $p_{tn}q_{rn}/p^t \cdot q^r$, can be regarded as predictors for the actual period r shares, s_{rn}. Since each share s_{tn} in the first set of terms is already weighted by its economic importance, there is no need for any further weighting of the first set of N squared terms in the summation to account for economic importance. The same analysis applies to the second set of N sum of squared terms; each term in the summation is already weighted by its economic importance.

If prices in period t are proportional to prices in period r (so that $p^t = \lambda_t p^r$ for some scalar $\lambda_t > 0$ or $p^r = \lambda_r p^t$ for some $\lambda_r > 0$), then it is easy to verify that $\Delta_{SP}(p^r, p^t, q^r, q^t)$ defined by (11.218) is equal to 0.

Now consider the implications on p^t and p^r if $\Delta_{SP}(p^r, p^t, q^r, q^t) = 0$. We need to consider a number of cases, depending on assumptions about the positivity of the prices and quantities in periods r and t. In all cases listed below, it is assumed that $p^r \cdot q^t > 0$ for $r = 1, \ldots, T$ and $t = 1, \ldots, T$.

Case (i): $\Delta_{SP}(p^r, p^t, q^r, q^t) = 0$ and $q^t >> 0_N$ *or* $q^r >> 0_N$; i.e., assume that all components of the period t *or* period r quantity vectors are positive. If $q^t >> 0_N$ and $\Delta_{SP}(p^r, p^t, q^r, q^t)$ defined by (11.218) is 0, then the first sum of squared terms, $\sum_{n=1}^{N} \left[s_{tn} - \left(p_{rn}q_{tn} / p^r \cdot q^t\right)\right]^2 = 0$, which implies that $p_{tn}q_{tn} = \left(p^t \cdot q^t / p^r \cdot q^t\right)p_{rn}q_{tn}$ which in turn implies that $p_{tn} = \left(p^t \cdot q^t / p^r \cdot q^t\right)p_{rn}$ since $q_{tn} > 0$ for $n = 1, \ldots, N$. Thus $p^t = \lambda_{tr} p^r$ where $\lambda_{tr} \equiv p^t \cdot q^t / p^r \cdot q^t > 0$ which implies that the period t price vector is proportional to the period r price vector. If $q^r >> 0_N$ and $\Delta_{SP}(p^r, p^t, q^r, q^t)$ is 0, then the second set of terms on the right hand side of (11.218) is equal to zero. Thus we must have $p_{rn} = \left(p^r \cdot q^r / p^t \cdot q^r\right)p_{tn}$ for $n = 1, \ldots, N$. Thus $p^r = \lambda_{rt} p^t$ where $\lambda_{rt} \equiv p^r \cdot q^r / p^t \cdot q^r > 0$ which in turn implies that the period r price vector is proportional to the period t price vector.

Case (ii): $\Delta_{SP}(p^r, p^t, q^r, q^t) = 0$ and $q^r + q^t >> 0_N$ so that each product is present in at least one of the two periods, periods r and t, whose prices are being compared. We further assume that there is at least one product n^* that is present in both periods being compared; i.e., there exists an n^* such that $q_{rn*} > 0$ and $q_{tn*} > 0$. Following the same type of argument that was pursued for Case (i) above, we find that our assumptions imply that $p_{tn} = \lambda_{tr} p_{rn}$ for n such that $q_{tn} > 0$ and $p_{rn} = \lambda_{rt} p_{tn}$ for n such that $q_{rn} > 0$. For products n^* that are present in both periods r and t, we have $p_{tn*} = \lambda_{tr} p_{rn*}$ and $p_{rn*} = \lambda_{rt} p_{tn*}$ and thus $\lambda_{tr} = 1/\lambda_{rt}$. These equalities imply that the period t price vector must be proportional to the period r price vector under our present assumptions.

Case (iii): Some products are not present in both periods r and t. This case can be reduced down to one of the previous cases for a new N^* that just includes the products that are present in at least one of periods r and t.

Using the above analysis, it can be seen that $\Delta_{SP}(p^r, p^t, q^r, q^t)$ equals 0 if and only if the period r and t price vectors are proportional. If the price vectors are not proportional, then $\Delta_{SP}(p^r, p^t, q^r, q^t)$ will be positive. A larger value for $\Delta_{SP}(p^r, p^t, q^r, q^t)$ indicates a bigger deviation from price proportionality. Thus $\Delta_{SP}(p^r, p^t, q^r, q^t)$ is a "reasonable" measure of bilateral relative price dissimilarity.

There are some aspects of the predicted price measure of relative price dissimilarity that require further discussion. When comparing the prices of periods r and t, suppose product 1 is present in period t but not present in period r. More precisely, suppose $q_{t1} > 0$ (and $p_{t1} > 0$) but $q_{r1} = 0$. What is the corresponding price for the missing product in period r; i.e., what exactly is p_{r1}? Suppose we set $p_{r1} = 0$. For simplicity, suppose further that prices and quantities for products 2 to N are the same in periods r and t, so that $p_{rn} = p_{tn}$ and $q_{rn} = q_{tn}$ for $n = 2, 3, \ldots, N$. Under these conditions, we find that $\Delta_{\mathrm{SP}}(p^r, p^t, q^r, q^t)$ is equal to the following sum of squared terms:

$$\begin{aligned}\Delta_{\mathrm{SP}}(p^r, p^t, q^r, q^t) \equiv & \sum_{n=1}^{N}\left[s_{tn} - \left(\frac{p_{rn}q_{tn}}{p^r \cdot q^t}\right)\right]^2 \\ & + \sum_{n=1}^{N}\left[s_{rn} - \left(\frac{p_{tn}q_{rn}}{p^t \cdot q^r}\right)\right]^2 \\ & = [s_{t1} - 0]^2 + \sum_{n=2}^{N}[s_{tn} - s_{rn}]^2 + \sum_{n=1}^{N}[s_{rn} - s_{rn}]^2 \\ & = s_{t1}^2 + \sum_{n=2}^{N}[s_{tn} - s_{rn}]^2 \\ & > 0 \end{aligned} \qquad (11.219)$$

where the inequality follows since under our assumptions, $s_{t1} > 0$. Thus even if all prices and quantities are the same for products that are present in both periods r and t, the dissimilarity measure defined by (11.218) will be positive as long as there are some products that are present in only one of the two periods being compared. Thus, if we set the prices for missing products equal to 0, then the predicted share measure of relative price dissimilarity will automatically register a positive measure; i.e., the measure will *penalize* a lack of matching of prices if we set the prices for missing products equal to 0.

Hill and Timmer were the first to point out the importance of having a measure of relative price dissimilarity that would penalize a lack of matching of the prices in the two periods being compared:

> In a survey of the literature on reliability measures, Rao and Timmer (2003) concluded that the main problem of existing measures, such as Hill's (1999a, 1999b) Paasche–Laspeyres spread and Diewert's (2002) class of relative price dissimilarity measures, is that they fail to make adjustments for gaps in the data. Rao and Timmer drew a distinction between statistical and index theoretic measures of reliability. The former takes a sampling perspective; bilateral comparisons based on a small number of matched product headings or a low coverage of total expenditure or production (averaged across the two countries) are deemed less reliable. In addition to the standard statistical arguments regarding small samples and a low coverage not being representative, little overlap in the product headings priced by the two countries implies that they are very different and, by implication, inherently difficult to compare. Index theoretic measures, in contrast, focus on the sensitivity of a bilateral comparison to the choice of price index formula. Most of the reliability measures proposed in the literature, including Hill's (1999a, 1999b) Paasche–Laspeyres spread and Diewert's (2002) class of relative price dissimilarity measures, are of this type. Although these measures perform well when there are few gaps in the data, they can generate highly misleading results when there are many gaps. This is because they fail to penalize bilateral comparisons made over a small number of matched headings.
>
> Robert J. Hill and Marcel P. Timmer (2006, 366)

The above considerations suggest that the predicted share measure of relative price dissimilarity could be used under two different sets of circumstances when there are missing prices:

- Use carry forward (or backward) prices or reservation prices for the missing prices and use the measure $\Delta_{SP}(p^r, p^t, q^r, q^t)$ defined by (11.218) to link the observations. With a complete set of prices for each period in hand, the usual bilateral Fisher index could be used as the linking index. This approach is consistent with the economic approach to index number theory.
- Do not estimate carry forward or reservation prices for the missing price observations (and set the prices of the missing products equal to 0) but still use $\Delta_{SP}(p^r, p^t, q^r, q^t)$ to link the observations. In this case, the *maximum overlap* bilateral Fisher index is used as the linking index for each pair of links chosen by the similarity linking method. This approach is more consistent with the stochastic approach to index number theory used by Hill and Timmer (2006).

Both strategies are illustrated for our empirical example in the Appendix.

Some additional properties of $\Delta_{SP}(p^r, p^t, q^r, q^t)$ are the following ones:

- *Symmetry*; i.e., $\Delta_{SP}(p^r, p^t, q^r, q^t) = \Delta_{SP}(p^t, p^r, q^t, q^r)$.
- *Invariance to changes in the units of measurement*.
- *Homogeneity of degree 0 in the components of q^r and q^t*, i.e., $\Delta_{SP}(p^r, p^t, \lambda_r q^r, \lambda_t q^t) = \Delta_{SP}(p^r, p^t, q^r, q^t)$ for all $\lambda_r > 0$ and $\lambda_t > 0$.
- *Homogeneity of degree 0 in the components of p^r and p^t*; i.e., $\Delta_{SP}(\lambda_r p^r, \lambda_t p^t, q^r, q^t) = \Delta_{SP}(p^r, p^t, q^r, q^t)$ for all $\lambda_r > 0$ and $\lambda_t > 0$.

The relative price dissimilarity indexes $\Delta_{SP}(p^r, p^t, q^r, q^t)$ defined by (11.218) can be used in place of the dissimilarity indexes $\Delta_{AL}(p^r, p^t, q^r, q^t)$ defined by (11.215) in section "Linking Based on Relative Price Similarity" in order to link together bilateral Fisher indexes. Thus set the new relative price similarity linked Fisher price index for period 1 equal to unity; i.e., set $P_{SP}^1 \equiv 1$. The period 2 index is set equal to $P_F(p^1, p^2, q^1, q^2)$, the Fisher index linking the period 2 prices to the period 1 prices.[199] Thus $P_{SP}^2 \equiv P_F(p^1, p^2, q^1, q^2)P_{SP}^1$. For period 3, evaluate the dissimilarity indexes $\Delta_{SP}(p^1, p^3, q^1, q^3)$ and $\Delta_{SP}(p^2, p^3, q^2, q^3)$ defined by (11.218). If $\Delta_{SP}(p^1, p^3, q^1, q^3)$ is the minimum of these two numbers, define $P_{SP}^3 \equiv P_F(p^1, p^3, q^1, q^3)P_{SP}^1$. If $\Delta_{SP}(p^2, p^3, q^2, q^3)$ is the minimum of these two numbers, define $P_{SP}^3 \equiv P_F(p^2, p^3, q^2, q^3)P_{SP}^2$. For period 4, evaluate the dissimilarity indexes $\Delta_{SP}(p^r, p^4, q^r, q^4)$ for $r = 1, 2, 3$. Let r^* be such that $\Delta_{SP}(p^{r*}, p^4, q^{r*}, q^4) = \min_r\{\Delta_{SP}(p^r, p^4, q^r, q^4); r = 1, 2, 3\}$.[200] Then define $P_{SP}^4 \equiv P_F(p^{r*}, p^4, q^{r*}, q^4)P_{SP}^{r*}$. Continue this process in the same manner; i.e., for period t, let r^* be such that $\Delta_{SP}(p^{r*}, p^t, q^{r*}, q^t) = \min_r\{\Delta_{SP}(p^r, p^t, q^r, q^t); r = 1, 2, \ldots, t-1\}$

[199] In the present context, it is not necessary to have all prices positive in computing the Fisher indexes. However, if the economic approach to index number theory is applied, then it is preferable to impute the missing prices. Missing quantities should be left at their 0 values using the economic approach.

[200] If the minimum occurs at more than one r, choose r^* to be the earliest of these minimizing periods.

and define $P_{SP}^t \equiv P_F(p^{r*}, p^t, q^{r*}, q^t) P_{SP}^{r*}$. Again, as in section "Linking Based on Relative Price Similarity", this procedure allows for the construction of similarity linked indexes in real time.

Using the scanner data listed in Appendix 1 which included reservation prices for missing products, the new similarity linked price indexes P_{SP}^t were calculated and compared to the price similarity linked price indexes P_{AL}^t that were defined in section "Linking Based on Relative Price Similarity". The new measure of relative price dissimilarity led to a different pattern of bilateral links: 7 of the 38 bilateral links changed when the dissimilarity measure was changed from $\Delta_{AL}(p^r, p^t, q^r, q^t)$ to $\Delta_{SP}(p^r, p^t, q^r, q^t)$. However, the price indexes generated by these alternative methods for linking observations were very similar: the sample averages for P_{AL}^t and P_{SP}^t were 0.97069 and 0.97109, respectively, and the correlation coefficient between the two indexes was 0.99681. Both indexes ended up at 0.9275. Thus, even though the two measures of price dissimilarity generated a different pattern of bilateral links, the underlying indexes P_{AL}^t and P_{SP}^t approximated each other very closely.

Both of the similarity linked price indexes P_{AL}^t and P_{SP}^t satisfy *a strong identity test*; i.e., if $p^r = p^t$, then $P_{AL}^r = P_{AL}^t$ and $P_{SP}^r = P_{SP}^t$. It is not necessary for q^r to equal q^t for this strong identity test to be satisfied. Thus these similarity linked indexes have an advantage over the corresponding GEKS and CCDI multilateral indexes in that in order to ensure that $P_{GEKS}^r = P_{GEKS}^t$ and $P_{CCDI}^r = P_{CCDI}^t$, we require that $p^r = p^t$ *and* $q^r = q^t$; i.e., we require that quantities be equal for the two periods as well as prices.

The above material can be adapted to measuring the *relative similarity of quantities* in place of prices. The incentive to use similarity of relative quantities is as follows: if the period r and t quantity vectors are proportional, then both the Laspeyres, Paasche and Fisher quantity indexes will be equal to this factor of quantity proportionality. In particular, if $q^r = q^t$, then the Laspeyres, Paasche, Fisher and any superlative quantity index will be equal to unity, without requiring p^t and p^r to be equal. Thus, when the quantity vectors are proportional, it makes sense to define the price indexes residually using the Product Test. Thus, define the following measure of *relative quantity similarity* between the quantity vectors for periods r and t as follows[201]:

$$\Delta_{SQ}(p^r, p^t, q^r, q^t) \equiv \sum_{n=1}^{N}\left[s_{tn} - \frac{p_{tn}q_{rn}}{p^t \cdot q^r}\right]^2 + \sum_{n=1}^{N}\left[s_{rn} - \frac{p_{rn}q_{tn}}{p^r \cdot q^t}\right]^2. \tag{11.220}$$

If the quantity vectors q^r and q^t are proportional to each other, then it is straightforward to verify that $\Delta_{SQ}(p^r, p^t, q^r, q^t) = 0$. On the other hand, if $\Delta_{SQ}(p^r, p^t, q^r, q^t) = 0$, then one can repeat Cases (i)–(iii) above, with prices and quantities interchanged, to show that q^r and q^t must be proportional to each other. Thus $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ equals 0 if and only if the period r and t quantity vectors are proportional. If the quantity vectors are not proportional, then $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ will be positive. A larger value for $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ indicates a bigger deviation from quantity proportionality. An advantage of the measure of dissimilarity defined by (11.220) is that it can deal with q_{tn} that are equal to 0.[202]

The new dissimilarity measure $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ can be used in place of $\Delta_{SP}(p^r, p^t, q^r, q^t)$ in order to construct a new pattern of bilateral Fisher price index links,[203] leading to a new series of price indexes, say P^t_{SQ} for $t = 1, \ldots, T$. The advantage in computing this sequence of price indexes is that they will satisfy the following *fixed basket test*: if $q^r = q^t \equiv q$ for $r < t$, then $P^t_{SQ}/P^r_{SQ} = p^t \cdot q / p^r \cdot q$. Note that this test does not require that $p^t = p^r$. Once the sequence of price indexes P^t_{SQ} has been constructed, corresponding quantity levels can be

[201] It can be seen that $\Delta_{SQ}(p^r, p^t, q^r, q^t) = \Delta_{SP}(q^r, q^t, p^r, p^t)$; i.e., the role of prices and quantities is interchanged in the above measure of price dissimilarity $\Delta_{SP}(p^r, p^t, q^r, q^t)$.

[202] If one takes the economic approach to index number theory and adopts the reservation price methodology due to Hicks (1940), then 0 prices can be avoided by using reservation prices or approximations to them such as inflation adjusted carry forward or backward prices. However, 0 quantities cannot be avoided so we need measures of price and quantity dissimilarity that can accommodate 0 prices and quantities in a sensible way.

[203] The implicit Fisher price index that is defined residually using the Product Test turns out to be equal to the usual Fisher price index that is defined directly as the geometric mean of the Laspeyres and Paasche price indexes.

defined as $Q^t_{SQ} \equiv p^t \cdot q^t \big/ P^t_{SQ}$ for $t = 1, \ldots, T$. The fixed basket test for price indexes translates into the following *strong identity test* for quantity indexes: if $q^r = q^t \equiv q$ for $r < t$, then $Q^t_{SQ} \big/ Q^r_{SQ} = 1$. Note that this test does not require that $p^r = p^t$. It can be seen that this is the advantage in using $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ as the dissimilarity measure in place of $\Delta_{SP}(p^r, p^t, q^r, q^t)$: if $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ is used, then the strong identity test for quantities will be satisfied by the resulting quantity indexes, Q^t_{SQ}. On the other hand if $\Delta_{SP}(p^r, p^t, q^r, q^t)$ is used as the measure of relative price dissimilarity, then the resulting price indexes P^t_{SP} will satisfy the strong identity test for prices.

It is possible to design a measure that combines relative price dissimilarity with relative quantity dissimilarity such that the resulting dissimilarity measure when used with Fisher price index bilateral links in the usual manner gives rise to a sequence of price indexes (relative to period 1) P^t_{SPQ} that will satisfy both the fixed basket test and the strong identity test for prices. Define the following index for *relative price and quantity dissimilarity* between periods r and t, $\Delta_{SPQ}(p^r, p^t, q^r, q^t)$, as follows[204]:

$$\Delta_{SPQ}(p^r, p^t, q^r, q^t) \equiv \min\{\Delta_{SP}(p^r, p^t, q^r, q^t), \Delta_{SQ}(p^r, p^t, q^r, q^t)\} \tag{11.221}$$

Thus if prices are equal to each other for periods r and t, then $\Delta_{SP}(p^r, p^t, q^r, q^t)$ and $\Delta_{SPQ}(p^r, p^t, q^r, q^t)$ will both equal 0 and our linking procedure will lead to equal price levels for periods r and t. On the other hand, if quantities are equal to each other for periods r and t, then $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ and $\Delta_{SPQ}(p^r, p^t, q^r, q^t)$ will both equal 0 and our linking procedure will lead to equal quantity levels for periods r

[204] This approach that combines measures of relative price dissimilarity with measures of relative quantity dissimilarity is due to Allen and Diewert (1981), Hill (2004) and Hill and Timmer (2006, 277). Hill and Timmer also noted that, usually, the relative price dissimilarity measure $\Delta_{SP}(p^r, p^t, q^r, q^t)$ will be smaller than the relative quantity dissimilarity measure $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ in which case the combined measure $\Delta_{SPQ}(p^r, p^t, q^r, q^t)$ reduces to the price measure $\Delta_{SP}(p^r, p^t, q^r, q^t)$. Allen and Diewert (1981) found this to be the case with their empirical example and we find the same to be true for our empirical example in the Appendix.

and t.[205] Denote the price indexes relative to period 1 generated using $\Delta_{SPQ}(p^r, p^t, q^r, q^t)$ as the measure of dissimilarity by P^t_{SPQ} for $t = 1, \ldots, T$. Call this method the *SPQ multilateral method*. Thus, the similarity linked indexes that are generated using the dissimilarity measure defined by (11.221) will lead to index levels that satisfy both a strong identity test for prices and a strong identity test for quantities. Thus if prices are identical in the two periods being compared ($p^r = p^t$), then the similarity linked price levels for periods r and t are equal *and* if quantities are identical in the two periods being compared ($q^r = q^t$), then the similarity linked quantity levels for periods r and t are equal. No of the other multilateral methods studied in this chapter have this very strong property. *This property rules out chain drift both in the price and quantity levels*.

Using the scanner data listed in Appendix 1, the new similarity linked price indexes that combine price and quantity similarity linking, P^t_{SPQ}, were calculated and compared to the price similarity linked price indexes P^t_{SP} that were defined in the beginning of this section. For our sample data set, it turned out that predicted share quantity dissimilarity was always greater than the corresponding measure of predicted share price dissimilarity for each pair of observations in our sample. Under these conditions, it can be seen that $\Delta_{SPQ}(p^r, p^t, q^r, q^t)$ will equal $\Delta_{SP}(p^r, p^t, q^r, q^t)$ for all periods r and t. Thus the same set of bilateral Fisher index links that were generated using $\Delta_{SP}(p^r, p^t, q^r, q^t)$ were also generated using $\Delta_{SPQ}(p^r, p^t, q^r, q^t)$ defined by (11.221) as the measure of dissimilarity. It turns out that it was always the case that $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ was much bigger than the corresponding $\Delta_{SP}(p^r, p^t, q^r, q^t)$; i.e., in all cases, *relative quantity dissimilarity was much bigger than the corresponding relative price dissimilarity*.[206]

In section A5 of the Appendix, some variations on the multilateral indexes P^t_{AT} and P^t_{SQ} are considered and evaluated using the price and quantity data for our empirical example. The indexes P^t_{ALM} and P^t_{SPM} use the same tables of dissimilarity measures that were used to define

[205] Thus a strong version of Walsh's multiperiod identity test will hold using this procedure; i.e., if $p^r = p^t$, then the period r and t price levels will coincide and if $q^r = q^t$, then the period r and t quantity levels will coincide. Note that these tests will hold no matter how large the number of observations T is.

[206] Allen and Diewert (1981) and Hill and Timmer (2006) found the same pattern for their empirical examples using their measures of price and quantity dissimilarity.

the bilateral links for the indexes P_{AT}^t and P_{SP}^t but instead of generating real-time indexes, the new *modified* indexes P_{ALM}^t and P_{SPM}^t use the observations for the first year of data in the sample to construct a *spanning tree* of comparisons; i.e., the Robert Hill (2001) methodology is used to construct the set of bilateral comparisons for all months in the first year such that the resulting set of bilateral comparisons minimizes the sum of the dissimilarity measures for the chosen bilateral links. Once the set of bilateral links for the first year has been determined, subsequent months are linked to previous months in real time. Thus, the bilateral links for P_{AT}^t and P_{ALM}^t to the index levels of previous months are the same for all months t beyond the first year. Similar comments apply to P_{SP}^t and P_{SPM}^t. It follows that the longer term trends in P_{AT}^t and P_{ALM}^t will be the same as will the trends in P_{SP}^t and P_{SPM}^t.[207]

The indexes P_{AT}^t, P_{SP}^t, P_{ALM}^t and P_{SPM}^t all use reservation prices for the prices of missing products. These reservation prices were estimated econometrically in an earlier study by Diewert and Feenstra (2017). It is not easy to estimate reservation prices. Moreover, reservation prices rely on the applicability of the economic approach to index number theory and many assumptions are required in order to implement this approach. Thus, many statistical agencies will want to avoid the use of estimated reservation prices when constructing their consumer price indexes. As was indicated in the discussion below Eq. (11.219), the predicted share measure of relative price dissimilarity $\Delta_{\text{SP}}(p^r, p^t, q^r, q^t)$ defined by (11.218) is well defined even if the prices for missing products are set equal to zero.[208] As was mentioned earlier in this section, it is possible to use $\Delta_{\text{SP}}(p^r, p^t, q^r, q^t)$ as a guide to linking the observations even if the prices of missing products are set equal to 0. We explain how alternative versions of P_{SP}^t and P_{SPM}^t can be produced when the price vectors p^t have 0 components for missing products in period t in the following paragraph.

In order to explain how the alternative version of P_{SP}^t (call it P_{SP}^{t*}) is computed, it is first necessary to calculate all possible *maximum overlap bilateral Fisher indexes* for every pair of observations in the sample. Denote the maximum overlap Fisher price index for period t

[207] For our empirical example, P_{AL}^t, P_{SP}^t, P_{ALM}^t and P_{SPM}^t all end up at the same level for the last month in our sample; see Table A.7 and Chart 9 in the Appendix.

[208] This is not the case for the Asymptotic Linear measure of relative price dissimilarity $\Delta_{\text{AL}}(p^r, p^t, q^r, q^t)$ defined by (11.215).

relative to the base period r as $P_{\text{FMO}}(p^r, p^t, q^r, q^t)$ for all observations r and t. When calculating $P_{\text{FMO}}(p^r, p^t, q^r, q^t)$, the usual inner products $p^r \cdot q^t = \sum_{n=1}^{N} p_{rn}q_{tn}$ that are used to construct the Fisher index between periods r and t are replaced by summations over n where n is restricted to products that are present in both periods r and t. These four restricted inner products can be constructed very efficiently using matrix operations. As noted above, the dissimilarity measure $\Delta_{\text{SP}}(p^r, p^t, q^r, q^t)$ defined by (11.218) is well defined even if the prices for missing products are set equal to zero. Set the maximum overlap similarity linked price index P_{SP}^{1*} for period 1 equal to unity; i.e., set $P_{\text{SP}}^{1*} \equiv 1$. The period 2 index P_{SP}^{2*} is set equal to $P_{\text{FMO}}(p^1, p^2, q^1, q^2)$, the maximum overlap Fisher index linking the period 2 prices to the period 1 prices. Thus $P_{\text{SP}}^{2*} \equiv P_{\text{FMO}}(p^1, p^2, q^1, q^2)P_{\text{SP}}^{1*}$. For period 3, evaluate the dissimilarity indexes $\Delta_{\text{SP}}(p^1, p^3, q^1, q^3)$ and $\Delta_{\text{SP}}(p^2, p^3, q^2, q^3)$ defined by (11.218). If $\Delta_{\text{SP}}(p^1, p^3, q^1, q^3)$ is the minimum of these two numbers, define $P_{\text{SP}}^{3*} \equiv P_{\text{FMO}}(p^1, p^3, q^1, q^3)P_{\text{SP}}^{1*}$. If $\Delta_{\text{SP}}(p^2, p^3, q^2, q^3)$ is the minimum of these two numbers, define $P_{\text{SP}}^{3*} \equiv P_{\text{FMO}}(p^2, p^3, q^2, q^3)P_{\text{SP}}^{2*}$. For period 4, evaluate the dissimilarity indexes $\Delta_{\text{SP}}(p^r, p^4, q^r, q^4)$ for $r = 1, 2, 3$. Let r° be such that $\Delta_{\text{SP}}(p^{r^\circ}, p^4, q^{r^\circ}, q^4) = \min_r\{\Delta_{\text{SP}}(p^r, p^4, q^r, q^4); r = 1, 2, 3\}$.[209] Then define $P_{\text{SP}}^{4*} \equiv P_{\text{FMO}}(p^{r^\circ}, p^4, q^{r^\circ}, q^4)P_{\text{SP}}^{r^\circ}$. Continue this process in the same manner; i.e., for period t, let r° be such that $\Delta_{\text{SP}}(p^{r^\circ}, p^t, q^{r^\circ}, q^t) = \min_r\{\Delta_{\text{SP}}(p^r, p^t, q^r, q^t); r = 1, 2, \ldots, t-1\}$ and define $P_{\text{SP}}^{t*} \equiv P_{\text{FMO}}(p^{r^\circ}, p^t, q^{r^\circ}, q^t)P_{\text{SP}}^{r^\circ}$. The procedure for constructing P_{SP}^{t*} is exactly the same as the procedure for constructing P_{SP}^{t} except that maximum overlap Fisher indexes are used in place of regular Fisher indexes defined over all products in order to implement the "best" set of bilateral links that are used to link all of the observations in the sample up to the current period t.[210]

209 If the minimum occurs at more than one r, choose r^* to be the earliest of these minimizing periods.

210 In addition to using P_{FMO} in place of P_F, the other difference in the two procedures is the use of 0 prices for unavailable products in place of reservation or carry forward prices when evaluating the dissimilarity measures $\Delta_{\text{SP}}(p^r, p^t, q^r, q^t)$. Thus the set of optimal bilateral links can change as we move from the P_{SP}^{t} indexes to their maximum overlap counterpart P_{SP}^{t*} indexes.

Recall the definition for the modified set of price levels P_{ALM}^{t} using the Asymptotic Linear measure of relative price dissimilarity, which were similar to the P_{AT}^{t} price levels except that a year of data on prices and quantities was used to form a set of bilateral links that minimizes the sum of the associated dissimilarity measures that link the observations for the first year. The same procedure can be used in the present context where the P_{SP}^{t*} can be replaced by the *Modified Predicted Share indexes*, P_{SPM}^{t*}.[211] For months t that follow after the first "training" year, the bilateral links are the same as the links used to calculate the Predicted Share indexes P_{SP}^{t*}.[212]

The maximum overlap fixed base Fisher indexes, $P_{FMO}(p^1, p^t, q^1, q^t) \equiv P_F^{t*}$, and the GEKS indexes P_{GEKS}^{t*} using maximum overlap Fisher indexes in place of regular Fisher indexes are listed in the Appendix and can be compared to their counterparts P_F^t and P_{GEKS}^t that used reservation prices for the missing products. See Table A.7 in section A5 of the Appendix for a listing of the following indexes: P_{AL}^t, P_{ALM}^t, P_{SP}^t, P_{SPM}^t, P_{SP}^{t*}, P_{SPM}^{t*}, P_{GEKS}^t, P_{GEKS}^{t*}, P_F^t, P_F^{t*}. The final level for these ten indexes after 3 years of data where the level in month 1 was 1.00000 was as follows: 0.92725, 0.92725, 0.92725, 0.92725, 0.92612, 0.92612, 0.94591, 0.94987, 0.95071 and 0.95610. Thus, the first four similarity linked indexes end up at the same price level, 0.92575, while the predicted share and modified predicted share indexes that used maximum overlap prices, P_{SP}^{t*} and P_{SPM}^{t*}, ended up at the same slightly higher price level, 0.92612. The two GEKS indexes (P_{GEKS}^t used reservation prices while P_{GEKS}^{t*} used maximum overlap Fisher links that did not depend on any imputed prices) ended up about 2 percentage points above the similarity linked indexes. Finally, the fixed base Fisher index that used reservation prices and the fixed base Fisher index that

[211] Note that we cannot construct P_{AL}^{t*} or P_{ALM}^{t*} in the present context where we have 0 prices for missing products because $\Delta_{AL}(p^r, p^t, q^r, q^t)$ is not well defined when some prices are equal to zero.

[212] It is straightforward to apply the predicted share methodology when we have zero prices and quantities for missing products to quantity indexes. Apply definition (11.221); i.e., define $\Delta_{SPQ}(p^r, p^t, q^r, q^t) = \min\{\Delta_{SP}(p^r, p^t, q^r, q^t), \Delta_{SQ}(p^r, p^t, q^r, q^t)\}$ as our new measure of relative price and quantity dissimilarity where 0 prices and quantities are allowed to appear in the price and quantity vectors. Using this measure of dissimilarity and maximum overlap Fisher price and quantity indexes leads to the price levels P_{SPQ}^{t*}. For our empirical example, it was the case that $\Delta_{SP}(p^r, p^t, q^r, q^t)$ was always less than $\Delta_{SQ}(p^r, p^t, q^r, q^t)$ so the P_{SPQ}^{t*} ended up being equal to the P_{SP}^{t*} for all t.

used maximum overlap bilateral links, P_F^t and P_F^{t*}, ended up about 3 percentage points above the similarity linked index levels. These results lead to two important (but tentative) conclusions:

- The similarity linked indexes considered in this section and the previous sections all generate approximately the same results and
- The similarity linked indexes appear to generate lower rates of overall price change than the fixed base Fisher or the GEKS indexes generate.

The first dot point is important one if it is consistent with other empirical investigations. Some statistical agencies may prefer to use inflation adjusted carry forward prices to replace missing prices while other agencies may not wish to use any form of an imputed price in their indexes. The results for our empirical example suggest that it may not matter very much which strategy is chosen, provided similarity linking of observations is used.

The Axiomatic Approach to Multilateral Price Levels

In this section, we will look at the axiomatic or test properties of the five major multilateral methods studied in previous sections. The multilateral methods are the GEKS, CCDI, GK, WTPD and SPQ (Price and Quantity Similarity Linking) methods. The price levels for period t for the five methods are defined by definitions (11.69) for p_{GEKS}^t (11.76) for p_{CCDI}^t (11.137) for p_{GK}^t (11.210) for p_{WTPD}^t and by (11.221) for P_{SPQ}^t.[213] We will look at the properties of these price level functions rather than at the corresponding price indexes.[214] Denote the period t price level function for generic multilateral method M as $p_M^t(p^1, \ldots, p^T; q^1, \ldots, q^T)$ for $t = 1, \ldots, T$. We will follow the example of Dalén (2001, 2017) and Zhang et al. (2019) in considering a *dynamic product universe*; i.e., we

[213] The price and quantity similarity linked price levels P_{SPQ}^t have been normalized to equal 1 in period 1. The other four sets of price levels have not been normalized.

[214] For earlier work on the axiomatic properties of multilateral price and quantity indexes, see Diewert (1988, 1999b) and Balk (2008). These earlier studies did not look at the properties of stand alone price level functions.

will allow for new products and disappearing products in the tests that follow. N is the total number of products that are in the aggregate over all T periods. If a product n is not available in period t, we set q_{tn} equal to 0. We will assume that the corresponding price p_{tn} is a positive Hicksian reservation price or a positive inflation adjusted carry forward or backward price. Thus for each period t, the price vector $p^t >> 0_N$ but the corresponding period t quantity vector satisfies only $q^t > 0_N$; i.e., the missing products in period t are assigned 0 values for the corresponding quantities.[215] It proves convenient to define the N by T matrices of prices and quantities as $P \equiv [p^1, \ldots, p^T]$ and $Q \equiv [q^1, \ldots, q^T]$. Thus the p^t and q^t are to be interpreted as column vectors of dimension N in the definitions of the matrices P and Q.

Consider the following nine tests for a system of generic multilateral price levels, $p_M^t(P, Q)$:

Test 1: *The strong identity test for prices*. If $p^r = p^t$, then $p_M^r(P, Q) = p_M^t(P, Q)$. Thus if prices are equal in periods r and t, then the corresponding price levels are equal even if the corresponding quantity vectors q^r and q^t are not equal.

Test 2: *The fixed basket test for prices* or *the strong identity test for quantities*.[216] If $q^r = q^t \equiv q$, then the price index for period t relative to period r is $p_M^t(P, Q)/p_M^r(P, Q)$ which is equal to $p^t q/p^r q$.[217]

Test 3: *Linear homogeneity test for prices*. Let $r \neq t$ and $\lambda > 0$. Then $p_M^t\left(p^1, \ldots, p^{t-1}, \lambda p^t, p^{t+1}, \ldots, p^T, Q\right)/p_M^r\left(p^1, \ldots, p^{t-1}, \lambda p^t, p^{t+1}, \ldots, p^T, Q\right) = \lambda p_M^t(P, Q)/p_M^r(P, Q)$. Thus if all prices in period t are multiplied by a common scalar factor λ, then the price level of period t relative to the price level of any other period r will increase by the multiplicative factor λ.

[215] It is necessary to have strictly positive prices in order to calculate the CCDI price levels. The remaining multilateral methods do not require strictly positive prices for all products and all periods to be well defined but our last test involves imputed prices for missing products. Thus we need to introduce these imputed prices at the outset of our axiomatic framework.

[216] The period t quantity level that matches up with the period t price level is $q_M^t(P, Q) \equiv p^t q^t/p_M^t(P, Q)$ for $t = 1, \ldots, T$. Test 2 translates into the *strong identity test for quantity levels*; i.e., if $q^r = q^t$, then $q_M^r(P, Q) = q_M^t(P, Q)$ even if the price vectors for the two periods p^r and p^t are not equal.

[217] Tests 1 and 2 are essentially versions of Tests 1 and 2 suggested by Zhang et al. (2019).

Test 4: *Homogeneity test for quantities*. Let $\lambda > 0$. Then $p_M^r(P, q^1, \ldots, q^{t-1}, \lambda q^t, q^{t+1}, \ldots, q^T) = p_M^r(P, Q)$ for $r = 1, \ldots, T$. Thus if all quantities in period t are multiplied by a common scalar factor λ, then the price level of any period r remains unchanged. This property holds for all $t = 1, \ldots, T$.

Test 5: *Invariance to changes in the units of measurement*. The price level functions $p_M^t(P, Q)$ for $t = 1, \ldots, T$ remain unchanged if the N commodities are measured in different units of measurement.

Test 6: *Invariance to changes in the ordering of the commodities*. The price level functions $p_M^t(P, Q)$ for $t = 1, \ldots, T$ remain unchanged if the ordering of the N commodities is changed.

Test 7: *Invariance to changes in the ordering of the time periods*. If the T time periods are reordered by some permutation of the first T integers, then the new price level functions are equal to the same permutation of the initial price level functions. This test is considered to be an important one in the context of making cross sectional comparisons of price levels across countries. In the country context, if this test is satisfied, then all countries are treated in a symmetric manner. It is not so clear whether this test is important in the time series context.

Test 8: *Responsiveness to Isolated Products Test*: If a product is available in only one period in the window of T periods, this test asks that the price level functions $p_M^t(P, Q)$ *respond to changes* in the prices of these isolated products; i.e., the test asks that the price level functions *are not constant* as the prices for isolated products change. This test is a variation of Test 5 suggested by Zhang, Johansen and Nygaard (2019), which was a bilateral version of this test.[218]

Test 9: *Responsiveness to Changes in Imputed Prices for Missing Products Test*: If there are missing products in one or more periods, then there will be imputed prices for these missing products according to our methodological framework. This test asks that the price level functions $p_M^t(P, Q)$ *respond to changes* in these imputed prices, i.e., the test asks that the price level functions *are not constant* as the imputed prices change. This test is essentially an extension of the previous Test 8. This test allows a price level to decline if new products enter the market place during the period and for consumer utility to increase as the number of available products

[218] This test was explicitly suggested by Claude Lamboray. Some care is needed in interpreting this test since the test framework assumes that there are imputed prices for the missing products.

increases. If this test is not satisfied, then the price levels will be subject to *new products bias*.[219] This is an important source of bias in a dynamic product universe.

It can be shown that GEKS and CCDI fail Tests 1 and 2, GK fails 1,4, 8 and 9, WTPD fails 1, 2,[220] 8 and 9 and SPQ fails 7. The above five multilateral methods pass the remaining Tests. Since Test 7 may not be so important in the time series context, it appears that the price and quantity similarity method of linking, the SPQ method, is "best" for the above tests. However, other reasonable tests could be considered in a more systematic exploration of the test approach to multilateral comparisons so our endorsement of the SPQ method is tentative at this point. Furthermore, the method needs to be tested on alternative data sets to see if "reasonable" indexes are generated by the method.

Summary of Results

Some of the more important results in each section of the chapter will be summarized here.

- If there are divergent trends in product prices, the Dutot index is likely to have an upward bias relative to the Jevons index; see section "Comparing CES Price Levels and Price Indexes".
- The Carli index has an upward bias relative to the Jevons index (unless all prices move proportionally over time in which case both indexes will capture the common trend). The same result holds for the weighted Carli (or Young) index relative to the corresponding weighted Jevons index; see section "Using Means of Order *r* to Aggregate Price Ratios".
- The useful relationship (11.41) implies that the Fisher index P_F^t will be slightly less than the corresponding fixed base Törnqvist

[219] On new goods bias, see Boskin et al. (1996), Nordhaus (1997), Diewert (1998) and the references in section "Time Product Dummy Regressions: The Case of Missing Observations" of Diewert (2021a).

[220] The Weighted Time Product Dummy price levels fail Test 2 if definitions (11.205) are used to define the period t price levels. This is the option that statistical agencies are using at present. However, The WTPD price levels P^{t**} n and the corresponding quantity levels Q^{t**} defined by (11.206) will satisfy Test 2. If all errors are equal to 0, Eqs. (11.205) and (11.206) will generate the same estimated price and quantity levels.

index P_T^t, provided that the products in scope for the index are highly substitutable and there are divergent trends in prices; see section "Relationships Between Some Share Weighted Price Indexes". Under these circumstances, the following inequalities between the Paasche, Geometric Paasche, Törnqvist, Geometric Laspeyres and Laspeyres indexes are likely to hold: $P_P^t < P_{\text{GP}}^t < P_T^t < P_{\text{GL}}^t < P_L^t$.

- The covariance identity (11.48) provides an exact relationship between the Jevons and Törnqvist indexes. Some conditions for equality and for divergence between these two indexes are provided at the end of section "Relationships Between the Jevons, Geometric Laspeyres, Geometric Paasche and Törnqvist Price Indexes".
- In section "Relationships between Superlative Fixed Base Indexes and Geometric Indexes That Use Average Annual Shares as Weights", a geometric index that uses annual expenditure sales of a previous year as weights, $P_{J\alpha}^t$, is defined and compared to the Törnqvist index, P_T^t. Eq. (11.62) provides an exact covariance decomposition of the difference between these two indexes. If the products are highly substitutable and there are divergent trends in prices, then it is likely that $P_T^t < P_{J\alpha}^t$.
- Section "To Chain or Not to Chain" derives an exact relationship (11.65) between the fixed base Törnqvist index, P_T^t, and its chained counterpart, P_{TCh}^t. This identity is used to show that it is likely that the chained index will "drift" below its fixed base counterpart if the products in scope are highly substitutable and prices are frequently heavily discounted. However, a numerical example shows that if quantities are slow to adjust to the lower prices, then upward chain drift can occur.
- Section "Relationships Between the Törnqvist Index and the GEKS and CCDI Multilateral Indexes" introduces two multilateral indexes, P_{GEKS}^t and P_{CCDI}^t. The exact identity (11.78) for the difference between P_{CCDI}^t and P_T^t is derived. This identity and the fact that P_F^t usually closely approximates P_T^t lead to the conclusion (11.79) that typically, P_F^t, P_T^t, P_{GEKS}^t and P_{CCDI}^t will approximate each other fairly closely.
- Section "Unit Value Price and Quantity Indexes" introduces the Unit Value price index P_{UV}^t and shows that if there are divergent trends in prices and the products are highly substitutable, it is likely that $P_{\text{UV}}^t < P_F^t$. However, this conclusion does not necessarily hold

if there are missing products in period 1. Section "Quality Adjusted Unit Value Price and Quantity Indexes" derives similar results for the Quality Adjusted Unit Value index, $P^t_{\text{UV}\alpha}$.

- Section "Relationships between Lowe and Fisher Indexes" looks at the relationship of the Lowe index, P^t_{Lo}, with other indexes. The Lowe index uses the quantities in a base year as weights in a fixed basket type index for months that follow the base year. In using annual weights of a previous year, this index is similar in spirit to the geometric index $P^t_{J\alpha}$ that was analyzed in section "Relationships Between Superlative Fixed Base Indexes and Geometric Indexes That Use Aerage Annual Shares as Weights". The covariance type identities (11.128) and (11.131) are used to suggest that it is likely that the Lowe index lies between the fixed base Paasche and Laspeyres indexes; i.e., it is likely that $P^t_P < P^t_{\text{Lo}} < P^t_L$. The identity (11.134) is used to suggest that the Lowe index is likely to have an upward bias relative to the fixed base Fisher index; i.e., it is likely that $P^t_F < P^t_{\text{Lo}}$. However, if there are missing products in the base year, then these inequalities do not necessarily hold.
- Section "Geary Khamis Multilateral Indexes" looks at an additional multilateral index, the Geary Khamis index, P^t_{GK} and shows that P^t_{GK} can be interpreted as a quality adjusted unit value index and hence using the analysis in section "Quality Adjusted Unit Value Price and Quantity Indexes", it is likely that the Geary Khamis price index has a downward bias relative to the Fisher index; i.e., it is likely that $P^t_{\text{GK}} < P^t_F$. However, if there are missing products in the first month of the sample, the above inequality will not necessarily hold.
- Sections "Time Product Dummy Regressions: The Case of No Missing Observations"–"Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations" look at special cases of Weighted Time Product Dummy indexes, P^t_{WTPD}. These sections show how different forms of weighting can generate very different indexes. Section "Weighted Time Product Dummy Regressions: The General Case" finally deals with the general case where there are T periods and missing products. The exact identity (11.214) is used to show that it is likely that P^t_{WTPD} is less than the corresponding fixed base Törnqvist Theil index, P^t_T, provided that the products are highly substitutable and there are no missing products in period 1. However, if there are missing products in period 1, the inequality can be reversed.

- It turns out that the following price indexes are not affected by reservation prices: the unit value price indexes P^t_{UV} and $P^t_{\text{UV}\alpha}$, the Geary Khamis indexes P^t_{GK}, and the Weighted Time Product Dummy indexes P^t_{WTPD}. Thus these indexes are not consistent with the economic approach to dealing with the problems associated with new and disappearing products and services.
- The final multilateral indexes were introduced in sections "Linking Based on Relative Price Similarity"–"Linking Based on Relative Price and Quantity Similarity". These indexes use bilateral Fisher price indexes to link the price and quantity data of the current period to a prior period. The prior period that is chosen minimizes a measure of relative price (or quantity) dissimilarity. Two main measures of relative price dissimilarity were studied: the AL or Asymptotic Linear measure $\Delta_{\text{AL}}(p^r, p^t, q^r, q^t)$ defined by (11.215) and the SP or Predicted Share measure $\Delta_{\text{SP}}(p^r, p^t, q^r, q^t)$ defined by (11.218). The role of prices and quantities can be interchanged in order to define the Predicted Share measure $\Delta_{\text{SQ}}(p^r, p^t, q^r, q^t)$ of relative quantity dissimilarity which can also be used to generate a set of bilateral Fisher price index links. Finally, the minimum of the $\Delta_{\text{SP}}(p^r, p^t, q^r, q^t)$ and $\Delta_{\text{SQ}}(p^r, p^t, q^r, q^t)$ measures can be taken to define the $\Delta_{\text{SPQ}}(p^r, p^t, q^r, q^t)$ measure of relative price and quantity dissimilarity; see definition (11.221). When observations are linked using this dissimilarity measure, the resulting price indexes satisfy both the identity test for prices and the corresponding identity price for quantities. Thus the SPQ method explained in section "Linking Based on Relative Price and Quantity Similarity" has attractive axiomatic properties as is explained in section "The Axiomatic Approach to Multilateral Price Levels". For our empirical example, relative quantity dissimilarity was always greater than relative price dissimilarity so the SP and SPQ price indexes were always identical.
- For our empirical example, the similarity linked price indexes P^t_{AL} and $P^t_{\text{SP}} = P^t_{\text{SPQ}}$ ended up about 2 percentage points below P^t_{GEKS} and P^t_{CCDI} which in turn finished about 1 percentage point below P^t_F and P^t_T and finally P^t_{GK} and P^t_{WTPD} finished about 1 percentage point above P^t_F and P^t_T; see Table A.6 and Chart 8 in the Appendix. All of these indexes captured the trend in product prices quite well. More research is required in order to determine whether these differences are significant and occur in other examples.

- It is difficult to calculate reservation prices using econometric techniques. Thus section "Inflation Adjusted Carry Forward and Backward Imputed Prices" looked at methods for replacing reservation prices by inflation adjusted carry forward and backward prices which are much easier to calculate.
- For our empirical example, the replacement of the reservation prices by inflation adjusted carried forward or backward prices did not make much difference to the multilateral indexes.[221] If the products in scope are highly substitutable for each other, then we expect that this invariance result will hold (approximately). However, if products with new characteristics are introduced, then we expect that the replacement of econometrically estimated reservation prices by carried forward and backward prices would probably lead to an index that has an upward bias.
- Finally, in section "Linking Based on Relative Price and Quantity Similarity", we introduced some similarity linked Fisher price indexes that did not require imputations for missing prices. These indexes used the Predicted Share measure of relative price dissimilarity which is well defined even if the prices of missing products are set equal to 0. The Fisher indexes that link pairs of observations that have the lowest measures of dissimilarity are maximum overlap Fisher indexes. For our empirical example, it turned out that these indexes were very close to their counterparts that used reservation prices for the missing prices. These no imputation indexes (denoted by P_{SP}^{t*} and P_{SPM}^{t*}) were calculated for our data set and listed in Table A.7 and plotted on Chart 9 in the Appendix.

Conceptually, the Price and Quantity Similarity linked indexes P_{SPQ}^{t} seem to be the most attractive solution for solving the chain drift problem since the strong identity tests for both prices and quantities will always be satisfied using this multilateral method.

The data used for the empirically constructed indexes are listed in the Appendix so that the listed indexes can be replicated and so that alternative solutions to the chain drift problem can be tested out by other statisticians and economists.

[221] See Table A.8 in the Appendix.

Conclusion

It is evident that there is no easy solution to the chain drift problem. The previous *Consumer Price Index* Manual tended to use the economic approach to index number theory as a guide to choosing between alternative index number formula; i.e., the *Manual* tended to recommend the use of a superlative index number formula as a target index. However, the existence of deeply discounted prices and the appearance and disappearance of products often lead to a substantial chain drift problem. Some of the difficulties stem from the fact that the *economic approach to index number theory* that dates back to Konüs (1924), Konüs and Byushgens (1926) and Diewert (1976) suffers from the following problems:

- The theory assumes that all purchased goods and services are consumed in the period under consideration. But in reality, when a good goes on sale at a deeply discounted price, the quantity purchased will not necessarily be consumed in the current period. If the good can be stored, it will decrease demand for the product in the subsequent period. The traditional economic approach to index number theory does not take the storage problem into account.
- Preferences over goods and services are assumed to be complete. In reality, consumers may not be aware of many new (and old) products; i.e., knowledge about products may be subject to a diffusion process.
- Our approach to the treatment of new and disappearing products uses the reservation price methodology due to Hicks (1940), which simply assumes that latent preferences for new products exist in the period before their introduction to the marketplace. Thus, the consumer is assumed to have unchanging preferences over all periods. Before a new product appears, the quantity of the product is set equal to 0 in the consumer's utility function. In reality, a new product may change the consumer's utility function. This makes the estimation of reservation prices very difficult if not impossible.
- Preferences are assumed to be the same across consumers so that they can be represented by a common linearly homogeneous utility function. Moreover, the preferences do not change over time. All of these assumptions are suspect.

In view of the fact that the assumptions of the economic approach to index number theory will not be satisfied precisely in the real world, we cannot rely entirely on this approach to guide advice to statistical agencies on how to deal with the chain drift problem. Thus, it would be useful to develop the test approach to multilateral index number theory in more detail.

So what exactly should statistical agencies do to deal with the chain drift problem when price and quantity are available for a stratum of the CPI? At our current state of knowledge, it seems that the following methods are acceptable:

- Rolling window GEKS and CCDI. Probably the "safest" method of linking the results of one window to the previous window is to use the mean method suggested by Ivancic et al. (2009) and Diewert and Fox (2020). This is the method used by the Australian Bureau of Statistics (2016). However, in the case of seasonal products that are not present in all periods of the year, rolling window GEKS and CCDI can be problematic and similarity linking is preferred.
- Bilateral linking based on Price (and Quantity) Similarity. This method seems very promising. It can be adapted to work in situations where there are imputed prices for missing products or in situations where imputed prices are not allowed. The resulting indexes are guaranteed to be free of chain drift.

If only price information is available and there are no missing prices, then the Jevons index is the best alternative to use (at least from the perspective of the test approach to index number theory).

If only price information is available and there are missing prices for some products for some periods, then the time product dummy method is probably the best index to use. This method reduces to the Jevons index if there are no missing prices.[222]

We conclude this section by noting some priorities for future research:

[222] However, in situations where there are many missing prices, it may be preferable to adapt the predicted share similarity linking methodology to the case where only price information is available. We will explore this possibility in another chapter which deals with strongly seasonal products.

- We need more studies on Price Similarity Linking, particularly in the context of strongly seasonal commodities.
- What is the "optimal" length of the time period for a CPI? Should statistical agencies produce weekly or daily CPIs in addition to monthly CPIs?[223]
- There is a conceptual problem in using retail outlet prices to construct a consumer price index, since tourists and governments also purchase consumer goods. It would be preferable to use the purchase data of domestic households in order to construct a CPI for residents of the country so that the welfare of residents in the country could be calculated. However, if we focus on individual households, the matching problems are substantial due to the infrequency of purchases of storable commodities. Thus, it will be necessary to aggregate over demographically and locationally similar households in order to calculate indexes that minimize the number of imputations. In the perhaps distant future, it will become possible in a cashless society to utilize the data of banks and credit card companies to track the universe of purchases of individual households and thus to construct more accurate consumer price indexes. However, this development will depend on whether credit and debit card consumer transactions are also coded for the type of purchase.
- A final problem that may require some research is how to combine elementary indexes that are constructed using scanner data with elementary indexes that use web scraped data on prices or data on prices collected by employees of the statistical agency. This does not seem to be a big conceptual problem: for strata that use scanner data, we end up with an aggregate price and quantity level for each stratum. For strata that use web scraped data or collector data, we end up with a stratum elementary price level for each period and consumer expenditure survey information will generate an estimated value of consumer expenditures for the stratum in question so the corresponding stratum quantity can be defined as expenditure divided by the elementary price level. Thus, the resulting CPI

[223] The problem with making the time period shorter is that the number of price matches will decline, leading to the need for more imputations. Also, the shorter the period, the more variance there will be in the unit value prices and the associated quantities, leading to indexes that will also have high variances. Thus, the shorter the period, the less accurate the resulting indexes will be.

will be of uneven quality (because the expenditure estimates will not be current for the web scraped categories) but it will probably be of better quality than a traditional price collector generated CPI. However, as mentioned above, another problem is that the scanner data will apply not only to expenditures of domestic households but also to tourists and governments. Thus, there is a need for more research on this topic of combining methods of price collection.

Acknowledgements The author thanks Corinne Becker-Vermeulen, Jan de Haan, Adam Gorajek, Robert Hill, Ronald Johnson, Claude Lamboray, Chris Li, Marshall Reinsdorf, Alicia Rambaldi, Prasada Rao, Chihiro Shimizu, Mick Silver, Zachary Weselake-George and Clément Yélou for helpful comments on this chapter.

References

Aizcorbe, A., Corrado, C., & Doms, M. (2000). *Constructing price and quantity indexes for high technology goods*. Industrial Output Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System.

Allen, R. C., & Diewert, W. E. (1981). Direct versus implicit superlative index number formulae. *Review of Economics and Statistics, 63*, 430–435.

Alterman, W. F., Diewert, W. E., & Feenstra, R. C. (1999). *International trade price indexes and seasonal commodities*. Bureau of Labor Statistics.

Armknecht, P., & Silver, M. (2014). Post-Laspeyres: The case for a new formula for compiling consumer price indexes. *Review of Income and Wealth, 60*(2), 225–244.

Arrow, K. J., Chenery, H. B., Minhas, B. S., & Solow, R. M. (1961). Capital-labor substitution and economic efficiency. *Review of Economics and Statistics, 63*, 225–250.

Aten, B., & Heston, A. (2009). Chaining methods for international real product and purchasing power comparisons: Issues and alternatives. In D. S. P. Rao (Ed.), *Purchasing power parities of currencies: Recent advances in methods and applications* (pp. 245–273). Edward Elgar.

Australian Bureau of Statistics. (2016, November 29). *Making greater use of transactions data to compile the consumer price index* (Information Paper 6401.0.60.003). ABS.

Balk, B. M. (1980). A method for constructing price indices for seasonal commodities. *Journal of the Royal Statistical Society, Series A, 143*, 68–75.

Balk, B. M. (1981). A simple method for constructing price indices for seasonal commodities. *Statistische Hefte, 22*(1), 1–8.
Balk, B. M. (1996). A comparison of ten methods for multilateral international price and volume comparisons. *Journal of Official Statistics, 12*, 199–222.
Balk, B. M. (2008). *Price and quantity index numbers*. Cambridge University Press.
Bortkiewicz, Lv. (1923). Zweck und Struktur einer Preisindexzahl. *Nordisk Statistisk Tidsskrift, 2*, 369–408.
Boskin, M. J., Dulberger, E., Gordon, R., Griliches, Z., & Jorgenson, D. (1996). *Toward a more accurate measure of the cost of living*. Final Report to the U.S. Senate Finance Committee. US Government Printing Office.
Carli, G.-R. (1804). Del valore e della proporzione de' metalli monetati. In *Scrittori classici italiani di economia politica* (Vol. 13, pp. 297–366). G.G. Destefanis (originally published in 1764).
Carruthers, A. G., Sellwood, D. J., & Ward, P. W. (1980). Recent developments in the retail prices index. *The Statistician, 29*, 1–32.
Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal, 92*, 73–86.
Chessa, A. G. (2016). A new methodology for processing scanner data in the Dutch CPI. *Eurona, 1*, 49–69.
Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *American Economic Review, 18*(1), 139–165.
Court, A. T. (1939). Hedonic price indexes with automotive examples. In *The dynamics of automobile demand* (pp. 99–117). General Motors Corporation.
Dalén, J. (1992). Computing elementary aggregates in the swedish consumer price index. *Journal of Official Statistics, 8*, 129–147.
Dalén, J. (2001, April 2–6). *Statistical targets for price indexes in dynamic universes*. Paper presented at the Sixth meeting of the Ottawa Group.
Dalén, J. (2017, May 10–12). *Unit values and aggregation in scanner data—Towards a best practice*. Paper presented at the 15th meeting of the Ottawa Group.
Davies, G. R. (1924). The problem of a standard index number formula. *Journal of the American Statistical Association, 19*, 180–188.
Davies, G. R. (1932). Index numbers in mathematical economics. *Journal of the American Statistical Association, 27*, 58–64.
de Haan, J. (2004a). *The time dummy index as a special case of the imputation Törnqvist index*. Paper presented at The Eighth Meeting of the International Working Group on Price Indices (the Ottawa Group).
de Haan, J. (2004b, December 12–14). *Estimating quality-adjusted unit value indices: Evidence from scanner data*. Paper presented at the Seventh EMG Workshop.

de Haan, J. (2008, December 10). *Reducing drift in chained superlative price indexes for highly disaggregated data*. Paper presented at the Economic Measurement Workshop, Centre for Applied Economic Research, University of New South Wales.

de Haan, J. (2010). Hedonic price indexes: A comparison of imputation, time dummy and re-pricing methods. *Jahrbücher für Nationökonomie und Statistik, 230*, 772–791.

de Haan, J. (2015, May 22). *Rolling year time dummy indexes and the choice of splicing method*. Room Document at the 14th meeting of the Ottawa Group. http://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s3room

de Haan, J., & Krsinich, F. (2014). Scanner data and the treatment of quality change in nonrevisable price indexes. *Journal of Business and Economic Statistics, 32*, 341–358.

de Haan, J., & Krsinich, F. (2018). Time dummy hedonic and quality-adjusted unit value indexes: Do they really differ? *Review of Income and Wealth, 64*(4), 757–776.

de Haan, J., & van der Grient, H. (2011). Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics, 161*, 36–46.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics, 4*, 114–145.

Diewert, W. E. (1978). Superlative index numbers and consistency in aggregation. *Econometrica, 46*, 883–900.

Diewert, W. E. (1988). Test approaches to international comparisons. In W. Eichhorn (Ed.), *Measurement in economics: Theory and applications of economic indices* (pp. 67–86). Physica-Verlag.

Diewert, W. E. (1992). Fisher ideal output, input and productivity indexes revisited. *Journal of Productivity Analysis, 3*, 211–248.

Diewert, W. E. (1995). *Axiomatic and economic approaches to elementary price indexes* (Discussion Paper No. 95-01). Department of Economics, University of British Columbia.

Diewert, W. E. (1998). Index number issues in the consumer price index. *Journal of Economic Perspectives, 12*(1), 47–58.

Diewert, W. E. (1999a). Index number approaches to seasonal adjustment. *Macroeconomic Dynamics, 3*, 1–21.

Diewert, W. E. (1999b). Axiomatic and economic approaches to international comparisons. In A. Heston & R. E. Lipsey (Eds.), *International and interarea comparisons of income, output and prices* (pp. 13–87). Studies in Income and Wealth, Volume 61. The University of Chicago Press.

Diewert, W. E. (2002). *Weighted country product dummy variable regressions and index number formulae* (Department of Economics, Discussion Paper 02-15). University of British Columbia.

Diewert, W. E. (2003a). Hedonic regressions: A consumer theory approach. In R. C. Feenstra & M. D. Shapiro (Eds.), *Scanner data and price indexes, studies in income and wealth* (Vol. 61, pp. 317–348). University of Chicago Press.

Diewert, W. E. (2003b, May 27–29). *Hedonic regressions: A review of some unresolved issues*. Paper presented at the Seventh Meeting of the Ottawa Group.

Diewert, W. E. (2004). *On the stochastic approach to linking the regions in the ICP* (Discussion Paper No. 04-16). Department of Economics, The University of British Columbia.

Diewert, W. E. (2005a). Weighted country product dummy variable regressions and index number formulae. *Review of Income and Wealth, 51*, 561–570.

Diewert, W. E. (2005b). Adjacent period dummy variable hedonic regressions and bilateral index number theory. *Annales D'Économie et de Statistique, 79*(80), 759–786.

Diewert, W. E. (2009). Similarity indexes and criteria for spatial linking. In D. S. P. Rao (Ed.), *Purchasing power parities of currencies: Recent advances in methods and applications* (pp. 183–216). Edward Elgar.

Diewert, W. E. (2012). *Consumer price statistics in the UK*. Office for National Statistics. http://www.ons.gov.uk/ons/guide-method/userguidance/prices/cpi-and-rpi/index.html

Diewert, W. E. (2013). Methods of aggregation above the basic heading level within regions. In *Measuring the real size of the world economy: The framework, methodology and results of the International Comparison Program—ICP* (pp. 121–167). The World Bank.

Diewert, W. E. (2014). *An empirical illustration of index construction using Israeli data on vegetables* (Discussion Paper 14-04). School of Economics, The University of British Columbia.

Diewert, W. E. (2018). *Scanner data, elementary price indexes and the chain drift problem* (Discussion Paper 18-06). Vancouver School of Economics, University of British Columbia.

Diewert, W. E. (2021a). The economic approach to index number theory. Chapter 5 in *Consumer price index theory*. International Monetary Fund. https://www.imf.org/en/Data/Statistics/cpi-manual

Diewert, W. E. (2021b). Elementary indexes. Chapter 6 in *Consumer price index theory*. International Monetary Fund. https://www.imf.org/en/Data/Statistics/cpi-manual

Diewert, W. E. (2021c). Quality adjustment methods. Chapter 8 in *Consumer price index theory*. International Monetary Fund. https://www.imf.org/en/Data/Statistics/cpi-manual

Diewert, W. E., & Feenstra, R. (2017). *Estimating the benefits and costs of new and disappearing products* (Discussion Paper 17-10). Vancouver School of Economics, University of British Columbia.

Diewert, W. E., Finkel, Y., & Artsev, Y. (2009a). Empirical evidence on the treatment of seasonal products: The Israeli experience. In W. E. Diewert, B. M. Balk, D. Fixler, K. J. Fox, & A. O. Nakamura (Eds.), *Price and productivity measurement: Volume 2: Seasonality* (pp. 53–78). Trafford Press.

Diewert, W. E., & Fox, K. J. (2020). Substitution bias in multilateral methods for CPI construction using scanner data. *Journal of Business and Economic Statistics*. https://doi.org/10.1080/07350015.2020.1816176

Diewert, W. E., Fox, K. J., & Schreyer, P. (2017). *The digital economy, new products and consumer welfare* (Discussion Paper 17-09). Vancouver School of Economics, The University of British Columbia.

Diewert, W. E., Huwiler, M., & Kohli, U. (2009). Retrospective price indices and substitution bias. *Swiss Journal of Economics and Statistics, 145*(20), 127–135.

Diewert, W. E., & von der Lippe, P. (2010). Notes on unit value index bias. *Journal of Economics and Statistics, 230*, 690–708.

Drobisch, M. W. (1871). Über die Berechnung der Veränderung der Waarenpreis und des Geldwertes. *Jahrbücher für Nationalökonomie und Statistik, 16*, 416–427.

Dutot, C. (1738). *Réflexions politiques sur les finances et le commerce* (Vol. 1). Les frères Vaillant et N. Prevost.

Eltetö, Ö., & Köves, P. (1964). On a problem of index number computation relating to international comparisons (in Hungarian). *Statisztikai Szemle, 42*, 507–518.

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *American Economic Review, 84*, 157–177.

Feenstra, R. C., & Shapiro, M. D. (2003). High-frequency substitution and the measurement of price indexes. In R. C. Feenstra, & M. D. Shapiro (Eds.), *Scanner Data and Price Indexes* (pp. 123–146). Studies in Income and Wealth Volume 64. The University of Chicago Press.

Fisher, I. (1922). *The making of index numbers*. Houghton-Mifflin.

Frisch, R. (1936). Annual survey of general economic theory: The problem of index numbers. *Econometrica, 4*, 1–39.

Geary, R. G. (1958). A note on comparisons of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society Series A, 121*, 97–99.

Gini, C. (1931). On the circular test of index numbers. *Metron, 9*(9), 3–24.

Gorajek, A. (2018). *Econometric perspectives on economic measurement* (Research Discussion Paper 2018-08). Reserve Bank of Australia, 65 Martin Pl, 2000.

Griliches, Z. (1971). Introduction: Hedonic price indexes revisited. In Z. Griliches (Ed.), *Price indexes and quality change* (pp. 3–15). Harvard University Press.

Handbury, J., Watanabe, T., & Weinstein, D. E. (2013). *How much do official price indexes tell us about inflation* (NBER Working Paper 19504). National Bureau of Economic Research.

Hardy, G. H., Littlewood, J. E., & Pólya, G. (1934). *Inequalities*. Cambridge University Press.

Hicks, J. R. (1940). The valuation of the social income. *Economica, 7*, 105–140.

Hill, R. J. (1997). A taxonomy of multilateral methods for making international comparisons of prices and quantities. *Review of Income and Wealth, 43*(1), 49–69.

Hill, R. J. (1999a). Comparing price levels across countries using minimum spanning trees. *The Review of Economics and Statistics, 81*, 135–142.

Hill, R. J. (1999b). International comparisons using spanning trees. In A. Heston & R. E. Lipsey (Eds.), *International and interarea comparisons of income, output and prices* (pp. 109–120). Studies in Income and Wealth Volume 61, NBER. The University of Chicago Press.

Hill, R. J. (2001). Measuring inflation and growth using spanning trees. *International Economic Review, 42*, 167–185.

Hill, R. J. (2004). Constructing price indexes across space and time: The case of the European Union. *American Economic Review, 94*, 1379–1410.

Hill, R. J. (2009). Comparing per capita income levels across countries using spanning trees: Robustness, prior restrictions, hybrids and hierarchies. In D.S. P. Rao (Ed.), *Purchasing power parities of currencies: Recent advances in methods and applications* (pp. 217–244). Edward Elgar.

Hill, R. J., Rao, D. S. P., Shankar, S., & Hajargasht, R. (2017, May 25–26). *Spatial chaining as a way of improving international comparisons of prices and real incomes*. Paper presented at the Meeting on the International Comparisons of Income, Prices and Production, Princeton University.

Hill, R. J., & Timmer, M. P. (2006). Standard errors as weights in multilateral price indexes. *Journal of Business and Economic Statistics, 24*(3), 366–377.

Hill, T. P. (1988). Recent developments in index number theory and practice. *OECD Economic Studies, 10*, 123–148.

Hill, T. P. (1993). Price and volume measures. In *System of national accounts 1993* (pp. 379–406). Eurostat, IMF, OECD, UN and World Bank.

Huang, N., Wimalaratne, W., & Pollard, B. (2015, May 20–22). *Choice of index number formula and the upper level substitution bias in the Canadian CPI*. Paper presented at the 14th Ottawa Group Meeting.

ILO/IMF/OECD/UNECE/Eurostat/The World Bank. (2004). *Consumer price index manual: Theory and practice* (P. Hill, Ed.). International Labour Office.

Inklaar, R., & Diewert, W. E. (2016). Measuring industry productivity and cross-country convergence. *Journal of Econometrics, 191*, 426–433.

Ivancic, L., Diewert, W. E., & Fox, K. J. (2009). *Scanner data, time aggregation and the construction of price indexes* (Discussion Paper 09-09). Department of Economics, University of British Columbia.

Ivancic, L., Diewert, W. E., & Fox, K. J. (2010). *Using a constant elasticity of substitution index to estimate a cost of living index: From theory to practice* (Australian School of Business Research Paper No. 2010). ECON 15, University of New South Wales.

Ivancic, L., Diewert, W. E., & Fox, K. J. (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics, 161*, 24–35.

Jevons, W. S. (1865). The variation of prices and the value of the currency since 1782. *Journal of the Statistical Society of London, 28*, 294–320.

Keynes, J. M. (1909). The method of index numbers with special reference to the measurement of general exchange value. Reprinted as in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes* (1983) (Vol. 11, pp. 49–156). Cambridge University Press.

Keynes, J. M. (1930). *Treatise on money* (Vol. 1). Macmillan.

Khamis, S. H. (1970). Properties and conditions for the existence of a new type of index number. *Sankhya B, 32*, 81–98.

Khamis, S. H. (1972). A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society Series A, 135*, 96–121.

Konüs, A. A. (1924). The problem of the true index of the cost of living. Translated in *Econometrica, 7*(1939), 10–29.

Konüs, A. A., & Byushgens, S. S. (1926). K probleme pokupatelnoi cili deneg. *Voprosi Konyunkturi, 2*, 151–172.

Kravis, I. B., Heston, A., & Summers, R. (1982). *World product and income: International comparisons of real gross product*. The Johns Hopkins University Press.

Krsinich, F. (2016). The FEWS index: Fixed effects with a window splice. *Journal of Official Statistics, 32*, 375–404.

Laspeyres, E. (1871). Die Berechnung einer mittleren Waarenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik, 16*, 296–314.

Leontief, W. (1936). Composite commodities and the problem of index numbers. *Econometrica, 4*, 39–59.

Lowe, J. (1823). *The present state of England in regard to agriculture, trade and finance* (2nd ed.). Longman, Hurst, Rees, Orme and Brown.

Marris, R. (1984). Comparing the incomes of nations: A critique of the international comparison project. *Journal of Economic Literature, 22*(1), 40–57.

Muellbauer, J. (1974). Household production theory, quality and the hedonic technique. *American Economic Review, 64*(6), 977–994.

Nordhaus, W. D. (1997). Do real output and real wage measures capture reality? The history of lighting suggests not. In T. F. Bresnahan & R. J. Gordon (Eds.), *The economics of new goods* (pp. 29–66). University of Chicago Press.

Office for National Statistics (ONS). (2020). *New index number methods in consumer price statistics*. Office for National Statistics.

Paasche, H. (1874). Über die Preisentwicklung der letzten Jahre nach den Hamburger Borsennotirungen. *Jahrbücher für Nationalökonomie und Statistik, 12*, 168–178.

Pakes, A. (2001). *A reconsideration of hedonic price indices with and application to PCs* (NBER Working Paper 8715). National Bureau of Economic Research.

Persons, W. M. (1921). Fisher's formula for index numbers. *Review of Economics and Statistics, 3*(5), 103–113.

Persons, W. M. (1928). The effect of correlation between weights and relatives in the construction of index numbers. *The Review of Economics and Statistics, 10*(2), 80–107.

Rao, D. S. P. (1995). *On the equivalence of the generalized Country-Product-Dummy (CPD) method and the Rao-system for multilateral comparisons* (Working Paper No. 5). Centre for International Comparisons, University of Pennsylvania.

Rao, D. S. P. (2004, June 30–July 3). *The Country-Product-Dummy method: A stochastic approach to the computation of purchasing power parities in the ICP*. Paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement.

Rao, D. S. P. (2005). On the equivalence of the weighted Country Product Dummy (CPD) method and the Rao system for multilateral price comparisons. *Review of Income and Wealth, 51*(4), 571–580.

Rao, D. S. P., & Hajargasht, G. (2016). Stochastic approach to computation of purchasing power parities in the international comparison program. *Journal of Econometrics, 191*(2), 414–425.

Rao, D. S. P., & Timmer, M. P. (2003). Purchasing power parities for industry comparisons using weighted Eltetö-Köves-Szulc (EKS) methods. *Review of Income and Wealth, 49*, 491–511.

Reinsdorf, M. (2007). Axiomatic price index theory. In M. Boumans (Ed.), *Measurement in economics: A handbook* (pp. 153–188). Elsevier.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*, 34–55.

Schlömilch, O. (1858). Über Mïttelgrössen verschiedener Ordnungen. *Zeitschrift für Mathematik und Physik, 3*, 308–310.

Sergeev, S. (2001, November 12–14). *Measures of the similarity of the country's price structures and their practical application*. Conference on the European Comparison Program, U. N. Statistical Commission. Economic Commission for Europe, Geneva.

Sergeev, S. (2009). Aggregation methods based on structural international prices. In D. S. P. Rao (Ed.), *Purchasing power parities of currencies: Recent advances in methods and applications* (pp. 274–297). Edward Elgar.

Shapiro, M. D., & Wilcox, D. W. (1997). Alternative strategies for aggregating prices in the CPI. *Federal Reserve Bank of St Louis Review, 79*(3), 113–125.

Silver, M. (2010). The wrongs and rights of unit value indices. *Review of Income and Wealth, 56*, S206–S223.

Silver, M. (2011). An index number formula problem: The aggregation of broadly comparable items. *Journal of Official Statistics, 27*(4), 1–17.

Silver, M., & Heravi, S. (2005). A failure in the measurement of inflation: Results from a hedonic and matched experiment using scanner data. *Journal of Business and Economic Statistics, 23*, 269–281.

Summers, R. (1973). International comparisons with incomplete data. *Review of Income and Wealth, 29*(1), 1–16.

Szulc, B. J. (1964). Indices for multiregional comparisons (in Polish). *Przeglad Statystyczny, 3*, 239–254.

Szulc, B. J. (1983). Linking price index numbers. In W. E. Diewert & C. Montmarquette (Eds.), *Price level measurement* (pp. 537–566). Statistics Canada.

Szulc, B. J. (1987). Price indices below the basic aggregation level. *Bulletin of Labour Statistics, 2*, 9–16.

Theil, H. (1967). *Economics and information theory*. North-Holland Publishing.

Törnqvist, L. (1936). The Bank of Finland's consumption price index. *Bank of Finland Monthly Bulletin, 10*, 1–8.

Törnqvist, L., & Törnqvist, E. (1937). Vilket är förhällandet mellan finska markens och svenska kronans köpkraft? *Ekonomiska Samfundets Tidskrift, 39*, 1–39. Reprinted as In *Collected Scientific Papers of Leo Törnqvist* (pp. 121–160). The Research Institute of the Finnish Economy, 1981.

Triplett, J. (1987). Hedonic functions and hedonic indexes. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave: A dictionary of economics* (Vol. 2, pp. 630–634). Stockton Press.

Triplett, J. (2004). *Handbook on hedonic indexes and quality adjustments in price indexes* (Directorate for Science, Technology and Industry, DSTI/DOC(2004)9). OECD.

Triplett, J. E., & McDonald, R. J. (1977). Assessing the quality error in output measures: The case of refrigerators. *The Review of Income and Wealth, 23*(2), 137–156.

University of Chicago. (2013). *Dominick's data manual*. Kilts Center for Marketing, Booth School of Business.

Vartia, Y. O. (1978). Fisher's five-tined fork and other quantum theories of index numbers. In W. Eichhorn, R. Henn, O. Opitz, & R. W. Shephard (Eds.), *Theory and applications of economic indices* (pp. 271–295). Physica-Verlag.

Vartia, Y., & Suoperä, A. (2018). *Contingently biased, permanently biased and excellent index numbers for complete micro data* (Unpublished Paper). http://www.stat.fi/static/media/uploads/meta_en/menetelmakehitystyo/contingently_biased_vartia_suopera_updated.pdf

von Auer, L. (2014). The generalized unit value index family. *Review of Income and Wealth, 60*, 843–861.

von Auer, L. (2019, May 8). *The nature of chain drift: Implications for scanner data price indices*. Paper presented at the 16th Meeting of the Ottawa Group.

Walsh, C. M. (1901). *The measurement of general exchange value*. Macmillan and Co.

Walsh, C. M. (1921). *The problem of estimation*. P.S. King & Son.

Walsh, C. M. (1921). Discussion. *Journal of the American Statistical Association, 17*, 537–544.

Whittaker, E. T., & Robinson, G. (1940). *The calculus of observations* (3rd ed.). Blackie & Sons.

Young, A. (1812). *An inquiry into the progressive value of money as marked by the price of agricultural products*. Macmillan.

Zhang, L.-C., Johansen, I., & Nygaard, R. (2019). Tests for price indices in a dynamic item universe. *Journal of Official Statistics, 35*(3), 683–697.

CHAPTER 12

The Stochastic Approach to International Price Comparisons

Gholamreza Hajargasht

Introduction

International comparisons of prices and real incomes are ubiquitous in cross-country economic analyses. They are required in construction of the Penn World Table, testing theories of trade and development, calculating the Human Development Index and other development indicators and in estimating global inequality and poverty levels, to name a few. The International Comparisons Program (ICP) led by the World Bank coordinates the efforts to provide such comparisons and disseminates the results on a regular basis. Results from the most recent round of the ICP for the benchmark year 2017 are available from World Bank (2020). One of the main outputs of ICP is Purchasing Power Parities (PPPs) of currencies. PPPs are used for converting national accounts aggregates such as Gross Domestic Product (GDP) to comparable measures across countries. The purpose of this paper is to provide an overview of the stochastic

G. Hajargasht (✉)
University of Queensland, Brisbane, QLD, Australia
e-mail: g.hajargasht@uq.edu.au

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*,
https://doi.org/10.1007/978-981-19-2023-3_12

approach to construction of PPPs along with their measures of reliability obtained as the main byproduct. The stochastic approach to index numbers complements the axiomatic and economic theoretic approaches to index numbers.[1]

The stochastic approach to price index numbers has a long history going back to Jevons, Edgeworth, Bowley and Fisher. This early literature was criticized[2] by Walsh (1924) and Keynes (1930) leading to its abandonment for several decades. The works of Clements and Izan (1981, 1987) and Balk (1980) in 1980s[3] as well as Selvanathan and Rao (1994), Clements et al. (2006) and the critical review of Diewert (1995) led to the revival of the stochastic approach.[4] Prasada Rao contributed to the resurgence of the stochastic approach. In particular, he has been a leading figure in the development of a second strand of literature on the stochastic approach to multilateral index numbers based on the Country-Product-Dummy (CPD) model. The first or the earlier strand of stochastic approach to multilateral comparisons can be found in Selvanathan and Rao (1994). Recent works of Rao (2004, 2005), Diewert (2004, 2005), Hajargasht and Rao (2010), Deaton (2012), Rao and Hajargasht (2016), and Hajargasht et al. (2019) have demonstrated the versatility of the CPD model to generate well-known index numbers and their reliability measures. Hajargasht and Rao (2021) have further developed the stochastic approach through the CPD and other methods to generate PPPs and their reliability measures. This paper reviews the stochastic approach to multilateral price index numbers focusing on this latter strand of the literature.

The stochastic approach to index numbers has several benefits. Firstly, it can produce reliability measures for computed indexes and PPPs from the ICP. Despite the importance of cross-country price and income comparisons and considerable efforts put into the compilation of PPPs, it is generally unknown how reliable the published PPPs are. Secondly,

[1] For reviews of the axiomatic and economic approaches to PPPs see Diewert (1999, 2013), Balk (1996, 2008), and Hill (1997).

[2] The early stochastic approaches were criticised mainly because they lead to unweighted indexes.

[3] See Aldrich (1992), Balk (2008), and Diewert (2020) for more on the history of the stochastic approach to index numbers.

[4] - Theil (1967) is a pre-1980s work that developed an appropriate weighted stochastic approach to the Tornqvist index.

the stochastic approach can be potentially used for testing and selection of indexes. There is a number of indexes with good axiomatic properties in the literature, but there is no single index that possesses all the properties, i.e. none is an ultimate winner. It may be possible to develop statistical tests based on stochastic approach to choose the "best" index for a particular application. But perhaps the most useful virtue of the stochastic approach is that it equips the practitioner with the regression toolkit which makes incorporation of various important elements such as item attributes, quality change and correlation across prices possible.

The structure of the paper is as follows: in section The ICP Framework, the basic framework of ICP for PPP compilation is briefly described. In section Notation and Main Aggregation Methods, some of the well-known index number methods for international price comparisons are reviewed. Section Stochastic Approach to Rao, IDB and GK Indexes develops the stochastic approach using the CPD model where the law of one price is used to derive several aggregation methods for the compilation of PPPs along with their reliability measures. In section Multilateral Comparisons at Basic Heading Level, we discuss the stochastic approach to the CPD model at the basic heading level. Section The Stochastic Approach to GEKS reviews the stochastic approach to Gini-Elteto-Koves-Szulc (GEKS) method for multilateral comparisons. The GEKS method is of special importance as it is the main aggregation method used in the ICP for aggregating price and expenditure data leading to estimates of PPPs.

The ICP Framework

ICP relies on price and national accounts expenditure data for a large number of goods and services provided by the participating countries. The process of PPP compilation in ICP is complex, details of the framework and the steps involved can be found in Deaton and Heston (2010), Rao (2013a) and World Bank (2013, 2020). For our purposes in this chapter, it suffices to think of the ICP framework as a two-stage process. Figure 12.1 sketches the stages involved in this process and lists the aggregation methods relevant to PPP computation at the two stages of the ICP.

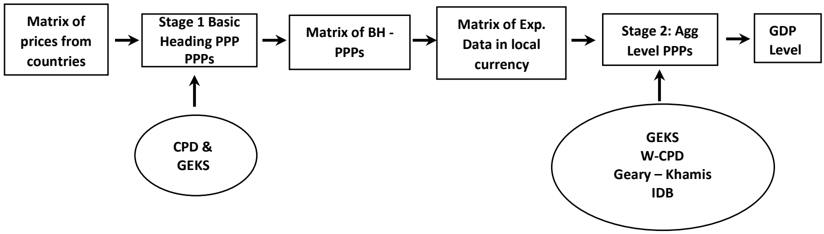

Fig. 12.1 A Schematic diagram of main steps in the ICP (*Source* Rao and Hajargasht [2016])

In the first stage, national prices of items within a given basic heading (BH)[5] are aggregated without weights using elementary indexes. The CPD and several variants of the Gini-Élteto-Köves-Szulc (GEKS) methods are used in the first stage.[6] The ICP at the World Bank uses the CPD method, whereas the OECD-Eurostat ICP uses variants of the GEKS method.

At the second stage, PPPs at the basic heading level are aggregated upwards using the available expenditure data to the GDP level or other major aggregates.[7] We will describe the GEKS, weighted CPD, Geary-Khamis (G-K) and Ikle-Dikhanov-Balk (IDB) methods used in the second stage of ICP in Sect. Notation and Main Aggregation Methods.[8] The ICP at the World Bank as well as at the OECD-Eurostat uses the GEKS method whereas the GK and IDB are used when additively consistent comparisons are desired.

[5] Basic heading level in the ICP is the lowest level of aggregation at which expenditure data are available.

[6] For reviews of these methods see the following sections or Rao (2013b).

[7] The ICP produces and publishes results for 26 aggregates and sub-aggregates. Major aggregates of interest are Consumption, Investment, Government Expenditure and GDP.

[8] See also Diewert (2013) for detailed descriptions of these methods.

Notation and Main Aggregation Methods

We consider the general case with M countries and N items.[9] Let p_{ij} represent the price of i-th item in j-th country and e_{ij} represents expenditure on i-th item in country j both expressed in national currency units.[10] Using this information, "*quantity*" can be implicitly defined as $q_{ij} = {e_{ij}}/{p_{ij}}$ and "*expenditure share*" as $w_{ij} = e_{ij} \Big/ \sum_{n=1}^{N} e_{nj}$. In order to describe some of the indexes, we also need to define $w^*_{ij} = w_{ij} \Big/ \sum_{m=1}^{M} w_{im}$ where $\sum_{m=1}^{M} w^*_{im} = 1$. Let PPP_j denote the purchasing power parity of currency of country j relative to a numeraire country. PPP_j shows the number of currency units of country j that have the same purchasing power, with respect to a basket of goods and services, as one unit of currency of a reference country. For example, if PPP for the currency of India with respect to one US dollar is equal to INR 15, then 15 Indian rupees in India have the same purchasing power as one US dollar in the United States. PPP between currencies of any two countries j and k, PPP_{jk}, can be obtained as the ratio PPP_j / PPP_k. The multilateral indexes used in the ICP ensure that the resulting PPPs are transitive and base invariant.[11]

Multilateral index numbers are often defined as a system of equations that determine both purchasing power parities of countries denoted by $PPP_j\ s$ and international average prices of items denoted by $P_i s$. The international average price of an item, as the name suggests, is an appropriately defined average of prices of the item across countries.

We are now ready to describe the principal aggregation methods (or multilateral indexes) for international price comparisons. Here we focus

[9] Items here refer to basic headings and therefore can be considered as composite commodities. In Sect. 5, we consider the problem of aggregating item level price data leading to PPPs at the basic heading level. The main difference is that no expenditure data are available at the item level.

[10] e_{ij} may be zero for some basic headings in some countries.

[11] See Rao (2013a) for a discussion of these properties.

on the weighted version of the methods.[12] These methods have also unweighted versions (see Table 12.1 for their formulas).

We start with Geary-Khamis (GK) system first proposed by Geary (1958), popularized by Khamis (1972) and was the main aggregation method in the ICP until 1985 (see Kravis et al., 1982, for a detailed discussion of the method). The GK system consists of the following system of $(M + N)$ equations.[13]

$$\begin{cases} \dfrac{1}{PPP_j} = \displaystyle\sum_{n=1}^{N} w_{nj} \dfrac{P_n}{p_{nj}} & j = 1, \ldots, M \\ P_i = \displaystyle\sum_{m=1}^{M} \dfrac{q_{im}}{\sum_{m=1}^{M} q_{im}} \dfrac{p_{im}}{PPP_m} & i = 1, \ldots, N \end{cases} \tag{12.1}$$

Note that here each PPP_j (i.e. the purchasing power of the j-th country) is defined as a weighted average of prices of the items in the j-th country deflated by corresponding international prices and each P_i is defined as weighted average of price of the i-th item across countries after they are converted into a common currency unit using PPPs of currencies of the corresponding countries. This is because prices in different countries are denominated in different currencies; therefore, it is necessary to convert them into a common currency unit before averaging.

A variety of other indexes have been proposed similarly by using alternative weights or alternative methods of averaging. For example, a system for international price comparisons proposed by Rao (1990) can be written as

$$\begin{cases} PPP_j = \displaystyle\prod_{n=1}^{N} \left(\dfrac{p_{nj}}{P_n} \right)^{w_{nj}} & j = 1, \ldots, M \\ P_i = \displaystyle\prod_{m=1}^{M} \left(\dfrac{p_{im}}{PPP_m} \right)^{w_{im}^{*}} & i = 1, \ldots, N \end{cases} \tag{12.2}$$

[12] See Diewert (1999) and Balk (2008) for axiomatic and economic properties of these methods.

[13] For conditions for existence of a unique positive solution to all of the indexes in this section, see Hajargasht and Rao (2019) and references cited therein.

Table 12.1 CPD Model specifications, moment conditions, unweighted and weighted indexes

Index	$r = \{r_{ij}\}$	*R matrix*	*Unweighted*	*Weights*	*Weighted*
Unweighted: *Jevons Geometric* ***Weighted*** *Rao System*	$r_{ij} = \ln p_{ij} - \ln P_i - \ln PPP_j$	$R' = D = \begin{bmatrix} i'_M \otimes I_N \\ I_M \otimes i'_M \end{bmatrix}$	$\ln \widehat{PPP}_j = \frac{1}{N}\sum_{n=1}^{N} \ln \frac{p_{nj}}{\hat{P}_n}$ $\ln \hat{P}_i = \frac{1}{M}\sum_{m=1}^{M} \ln \frac{p_{im}}{\widehat{PPP}_m}$	$W(w) = Diag\,(w_{ij})$	$\ln \widehat{PPP}_j = \sum_{n=1}^{N} w_{nj} \ln \frac{p_{nj}}{\hat{P}_n}$ $\ln \hat{P}_i = \sum_{m=1}^{M} w^*_{im} \ln \frac{p_{im}}{\widehat{PPP}_m}$
Arithmetic	$r_{ij} = \frac{p_{ij}}{P_i PPP_j} - 1$	$R' = D =$ $Diag\begin{pmatrix} -1/P \\ -1/PPP \end{pmatrix} \begin{bmatrix} i^i_M \otimes I_N \\ I_M \otimes i^i_N \end{bmatrix}$	$\widehat{PPP}_j = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{p_{nj}}{\widehat{P}_n}\right)$ $\widehat{P}_i = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{p_{im}}{\widehat{PPP}_m}\right)$	$W(w) = Diag(w_{ij})$	$\widehat{PPP}_j = \sum_{n=1}^{N} w_{nj} \frac{p_{nj}}{\hat{P}_n}$ $\hat{P}_i = \sum_{j=1}^{M} w^*_{im} \frac{p_{im}}{\widehat{PPP}_m}$
Unweighted: *Harmonic* ***Weighted:*** *Iklé*	$r_{ij} = \frac{P_i PPP_j}{p_{ij}} - 1$	$R' = D =$ $Diag\begin{pmatrix} 1/P \\ 1/PPP \end{pmatrix} \begin{bmatrix} i'_M \otimes I_M \\ I_M \otimes i'_N \end{bmatrix}$	$\frac{1}{\widehat{PPP}_j} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\widehat{P}_n}{p_{nj}}\right)$ $\frac{1}{\widehat{P}_i} =$ $\frac{1}{M}\sum_{m=1}^{M}\frac{\widehat{PPP}_m}{p_{im}}$	$W(w) = Diag(w_{ij})$	$\frac{1}{\widehat{PPP}_j} = \sum_{n=1}^{N} w_{nj} \frac{\hat{P}_n}{p_{nj}}$ $\frac{1}{\widehat{P}_i} = \sum_{m=1}^{M} w^*_{im} \frac{\widehat{PPP}_m}{p_{im}}$
Unweighted *Dutot* ***Weighted*** *GK*	$r_{ij} = \frac{p_{ij}}{P_i PPP_j} - 1$	$R' =$ $Diag\begin{pmatrix} -1/P \\ -1/PPP \end{pmatrix} \begin{bmatrix} i'_M \otimes I_N \\ I_M \otimes P'_N \end{bmatrix}$ $D =$ $Diag\begin{pmatrix} -1/P \\ -1/PPP \end{pmatrix} \begin{bmatrix} i'_M \otimes I_N \\ I_M \otimes P'_N \end{bmatrix}$	$\widehat{PPP}_j = \sum_{n=1}^{N} p_{nj} \Big/ \sum_{n=1}^{N} \widehat{P}_n$ $\widehat{P}_i = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{p_{im}}{\widehat{PPP}_m}\right)$	$\mathbf{W}(\boldsymbol{q}) = Diag\,(q_{ij})$	$\widehat{PPP}_j = \sum_{n=1}^{N} p_{nj} q_{nj} \Big/ \sum_{n=1}^{N} \hat{P}_n q_{nj}$ $\hat{P}_i = \sum_{m=1}^{M}\left(\frac{p_{im} q_{im}}{\widehat{PPP}_m}\right) \Big/ \sum_{m=1}^{M} q_{im}$

Adopted from Rao and Hajargasht (2016)

Diag represents a diagonal matrix in the arguments; $\mathbf{i}_M$ is $(M \times 1)$ column vector of ones; $\mathbf{I}_N$ is an identity matrix of size N; $W(\cdot)$ is the weight matrix. $1/P = \left(1/P_1 \cdots 1/P_N\right)'$; $1/\mathbf{PPP} = \left(1/PPP_1 \cdots 1/PPP_M\right)'$; $q_{ij} = e_{ij}/p_{ij}$; $w_{ij} = p_{ij}q_{ij} \Big/ \sum_{n=1}^{N} p_{nj}q_{nj}$; and $w^*_{ij} = w_{ij} \Big/ \sum_{m=1}^{M} w_{im}$

where w_{nj} s and w^*_{im} s are defined earlier in this section. As Rao (1995, 2005) has shown and we see later, this system can be obtained as a weighted least square estimator of parameters in the country-product-dummy (CPD) method.[14]

In the IDB system proposed by Iklé (1972) with its analytical properties discussed in Dikhanov (1994) andBalk (1996),[15] expenditure share weights are used along with harmonic averages as[16]

$$\begin{cases} \dfrac{1}{PPP_j} = \displaystyle\sum_{n=1}^{N}\left(w_{nj}\frac{P_n}{p_{nj}}\right) & j = 1, \ldots, M \\ \dfrac{1}{P_i} = \displaystyle\sum_{m=1}^{M}\left(w^*_{im}\frac{PPP_m}{p_{im}}\right) & i = 1, \ldots, N \end{cases} \tag{12.3}$$

Note that in the Rao system (12.2), PPPs and international average prices are geometric averages of national prices converted into a common currency using p_{ij}/PPP_j while in IDB system in (12.3) harmonic averages of the converted national prices are used in a similar manner. The GK and IDB methods are additive indexes, i.e. the real output of each country can be expressed as sum of the value of the country's individual consumption when each quantity is valued using a set of international prices which are constant across countries. This feature of an additive method is convenient for users and so for the purpose of analysing expenditure structures across countries. IDB has been found to be a better additive method than the GK method since it is less prone to the so-called Gerschenkron effect[17] (see, e.g., Dikhanov, 2020).

A lesser-known but still related system is the arithmetic index of Hajargasht and Rao (2010) where PPPs and Ps are arithmetic averages of price

[14] Details of this and other related results can be found in Rao (2009) and Rao and Hajargasht (2016).

[15] The system described in Iklé (1972) was difficult to follow and was ignored until the work of Dikhanov (1994) and Balk (1996) who provided an alternative formulation that is easy to understand and connects to other known index number systems.

[16] For conditions for existence of a unique positive solution to these systems, see Hajargasht and Rao (2019) and reference cited therein.

[17] The systematic overestimation of real GDP of poorer countries relative to the richer countries is referred to as the Gerschenkron effect.

ratios:

$$\begin{cases} PPP_j = \sum_{n=1}^{N} \left(w_{nj} \dfrac{p_{nj}}{P_n} \right) & j = 1, \ldots, M \\ P_i = \sum_{m=1}^{M} \left(w^*_{im} \dfrac{p_{im}}{PPP_m} \right) & i = 1, \ldots, N \end{cases} \tag{12.4}$$

The main method of aggregation used in ICP and arguably the method with best properties is the GEKS method due to Gini (1931), Eltetö and Köves (1964) and Szulc (1964).[18] The way GEKS is defined and motivated and its relevant stochastic approach is somewhat different. It uses the Fisher (or Tornqvist) binary index as building blocks for transitive multilateral comparisons.[19] The GEKS index is defined as follows:

$$PPP_j^{GEKS} = \prod_{l=1}^{M} \left[F_{jl} \cdot F_{lM} \right]^{1/M} \tag{12.5}$$

where PPP_j^{GEKS} denotes the GEKS purchasing power parity with respect to the base country and F_{jk} denotes the binary Fisher index between country j and k. Given the importance of the GEKS index, Sect. 6 is devoted to a more detailed analysis of this index and its relevant stochastic approach.

Stochastic Approach to Rao, IDB and GK Indexes[20]

The stochastic approach to computation of PPPs[21] can be based on what is known as the law of one price which postulates that the observed price, p_{ij}, of a commodity is approximately the product of its international price, P_i, and the purchasing power parity of the currency of the country j,

[18] There are also a number of less commonly used aggregation methods. The reader is referred to e.g. Balk (2008) for a discussion of some of these methods.

[19] See Rao (2009) for further details and a critique and generalization of the GEKS method for international comparisons.

[20] This Section draws heavily from Rao and Hajargasht (2016).

[21] One of the first studies that tried to compute standard error for PPPs is Kravis et al. (1975, pp. 77–79) using a Monte Carlo method. See Hajargasht and Rao (2021) for further explanations.

PPP_j. This can be written in its multiplicative form as

$$p_{ij} = P_i \cdot PPP_j \cdot \varepsilon_{ij} \tag{12.6}$$

where we have data on p_{ij}'s and the aim is to estimate P_i s and PPP_j s. ε_{ij} is an error term with mean equal to one. This equation is often written in its logarithmic form as $\ln p_{ij} = \gamma_i + \delta_j + \varepsilon^*_{ij}$ with $\delta_j = \ln PPP_j$, $\gamma_i = \ln P_i$ and $E(\varepsilon^*_{ij}) = 0$ The resulting regression is often estimated by introducing dummy variables for countries and items and therefore it is known as the Country-Product-Dummy (CPD) method. Rao (1995) was the first to note that the weighted least square estimates from the logarithmic version of CPD method leads to the geometric index (12.2).[22] Rao (2004, 2005), Diewert (2005), Deaton (2012), Hajargasht and Rao (2010), and Rao and Hajargasht (2016) further elaborated and extended the CPD model.

The stochastic approach based on the CPD model has several attractive features: (i) it enables us to incorporate data on characteristics such as information on outlets and geographical locations using suitably defined dummy variables (see, e.g., Hill & Syed, 2015 or Montero et al., 2020); (ii) weights reflecting importance of the priced items can be incorporated in a meaningful way[23]; and (iii) it allows incorporating heteroskedasticity as well as spatial dependence in the disturbance term.

Rao and Hajargasht (2016) have shown that, alternative ways of appending the error term in law of one price can lead to well-known aggregation methods. They show that model $\ln p_{ij} = \ln P_i + \ln PPP_j + \varepsilon^*_{ij}$ with $E(\varepsilon^*_{ij}) = 0$ leads to the geometric index (12.2); model $p_{ij} = P_i PPP_j \varepsilon_{ij}$ with $E(\varepsilon_{ij}) = 1$ leads to arithmetic index (12.4) and model $\frac{1}{p_{ij}} = \frac{1}{P_i PPP_j}\varepsilon_{ij}$ with $E(\varepsilon_{ij}) = 1$ leads to the harmonic index (12.3). Their analysis can be unified and extended by writing a generalized version of the law of one price as follows[24]:

$$p^{\rho}_{ij} = P^{\rho}_i PPP^{\rho}_j \varepsilon_{ij} \ with \ E(\varepsilon_{ij}) = 1 \tag{12.7}$$

[22] Summers (1973) used CPD as a method for filling missing price data in the context of international comparisons.

[23] In 2011 ICP, the general recommendation was to use 3:1 weights for products considered important.

[24] Gorajek (2018) has also developed a generalized stochastic model that can generate a variety of bilateral and multilateral indexes.

where ρ is known (i.e. it is not estimated); it can be any real number where, as we see later, values of 1, -1 and 0 are of particular interest leading to arithmetic, harmonic and geometric indexes, respectively. For $\rho = 0$, we write the model as

$$\ln p_{ij} = \ln P_i + \ln PPP_j + \varepsilon^*_{ij} \text{ with } E\left(\varepsilon^*_{ij}\right) = 0.$$

In general, variance of the vector of error terms $\varepsilon = [\varepsilon_{11} \cdots \varepsilon_{NM}]'$ can be specified as $Var(\varepsilon) = \Omega$.

Note that PPP_j s are identified up to a constant of proportionality. Therefore, a normalization is needed. It is common to assume $PPP_M = 1$ which means that PPP for the base country M[usually USA] is assumed to be one.

Unweighted Estimation: To estimate the model, rewrite (12.7) as

$$\left(p^{\rho}_{ij} \Big/ P^{\rho}_i PPP^{\rho}_j\right) - 1 = \varepsilon^*_{ij} \tag{12.8}$$

This is in the form of a non-linear non-additive regression model which can be estimated as described below.[25] Consider the regression model

$$r_{ij}(p_{ij}, \boldsymbol{P}, \boldsymbol{PPP}) = \varepsilon^*_{ij} \tag{12.9}$$

where $r_{ij} = \left(p^{\rho}_{ij} \Big/ P^{\rho}_i PPP^{\rho}_j\right) - 1$, $\mathbf{P} = [P_1 \ldots, P_N]'$ and $\mathbf{PPP} = [PPP_1, \ldots, PPP_{M-1}]'$.

This generalized model is in the form of a non-additive non-linear regression model[26] where the details of its estimation procedure can be found in Rao and Hajargasht (2016). There are N parameters representing international prices and $M-1$ parameters representing PPPs. We denote these $(N + M-1)$ parameters by vector $[\mathbf{P}\ \mathbf{PPP}]$. One approach to estimate model (12.9) is to use the method of moments with the following moment conditions:

$$E(\mathbf{R}'\varepsilon^*) = \mathbf{0} \tag{12.10}$$

[25] See Rao and Hajargasht (2016) and references cited there in for more details. The model (12.7) can also be estimated using a weighted maximum likelihood approach if a gamma distribution is assumed for ε_{ij}. Hajargasht and Rao (2010) show that this leads to the geometric system (12.2) if $\rho = 0$, to IDB system (12.3) if $\rho = -1$ and to arithmetic system (12.4) if $\rho = 1$.

[26] Applying least squares in such cases does not provide a consistent estimator.

where $\mathbf{R}$ is an $NM \times (N + M - 1)$ appropriately defined matrix of "instruments". By construction there are as many moment conditions as parameters, therefore, a method of moment (MoM) estimator can be obtained by solving the following sample moment conditions

$$\mathbf{R}(\mathbf{p}, \hat{\mathbf{P}}, \widehat{\mathbf{PPP}})' \mathbf{r}(\mathbf{p}, \hat{\mathbf{P}}, \widehat{\mathbf{PPP}}) = 0 \tag{12.11}$$

where r is a vector containing r_{ij} s. The resulting estimator is asymptotically normal with covariance matrix:

$$Var\left(\left[\hat{\mathbf{P}}\widehat{\mathbf{PPP}}\right]\right) = \left[\hat{\mathbf{D}}'\hat{\mathbf{R}}\right]^{-1} \hat{\mathbf{R}}'\hat{\Omega}\hat{\mathbf{R}}\left[\hat{\mathbf{R}}'\hat{\mathbf{D}}\right]^{-1} \tag{12.12}$$

with $\hat{\mathbf{D}} = E\left(\frac{\partial \mathrm{r}(\mathrm{p}, \hat{\mathbf{P}}, \widehat{\mathbf{PPP}})'}{\partial[\hat{\mathbf{P}}, \widehat{\mathbf{PPP}}]}\right)$, $\hat{\mathbf{R}} = \mathbf{R}(\mathbf{p}, \hat{\mathbf{P}}, \widehat{\mathbf{PPP}})$. $\hat{\boldsymbol{\Omega}}$ depends on the specification of the covariance matrix.[27] For example, if $\Omega = \sigma^2\mathbf{I}$ we can estimate $\widehat{\sigma^2} = \hat{\boldsymbol{r}}'\hat{\boldsymbol{r}} / NM$ and if $\Omega = diag(\sigma_{ij}^2)$ one can use White estimator $\widehat{\sigma_{ij}^2} = \hat{r}_{ij}^2$.

One issue with the above estimation procedure is the specification of $\mathbf{R}(\mathbf{p}, \mathbf{P}, \mathbf{PPP})$. Different choices are possible. However, it has been shown that the most efficient choice is (see, e.g., Wooldridge, 2010):

$$\mathbf{R}(\mathbf{p}, \mathbf{P}, \mathbf{PPP})_{\text{optimal}} = E\left[\frac{\partial \mathrm{r}(\mathbf{p}, \mathbf{P}, \mathbf{PPP})'}{\partial[\mathbf{P}, \mathbf{PPP}]}\right] \tag{12.13}$$

For general non-linear non-additive regressions, the expectation in the right-hand side of (12.13) cannot be derived analytically unless very strong distributional assumptions are made. Fortunately, for the type of models considered here, optimal $\mathbf{R}$ is tractable without making such assumptions. For the unweighted model (12.8) the optimal $\mathbf{R}$ and solution to (12.11) are given in (12.17), (12.18) and (12.19). Note that in case of an optimal $\mathbf{R}$, formula for the variance simplifies to

$$Var(\hat{\mathbf{P}}, \widehat{\mathbf{PPP}}) = [\hat{\mathbf{R}}' \hat{\mathbf{R}}]^{-1} \hat{\mathbf{R}}' \hat{\Omega}\hat{\mathbf{R}}\left[\hat{\mathbf{R}}'\hat{\mathbf{R}}\right]^{-1} \tag{12.14}$$

[27] Correlation across items and regions in the CPD model has been found to have important implications on the computed reliability measures. For less restrictive forms of the covariance matrix where correlations are allowed see Hajargasht and Rao (2021).

Use of the method of moments (12.11) along with different choices of r_{ij} s and $\mathbf{R}$ leads to different unweighted index numbers. Rao and Hajargasht (2016) show that with the right choices, it is possible to derive the unweighted geometric Jevons, arithmetic and harmonic averages of price relatives, and also the Dutot index (see Table 12.1 for further details).

Weighted Estimation: Since we are also interested in weighted index numbers with weights that reflect the relative importance of different basic headings, we appeal to weighted method of moments. Let $\mathbf{W}$ be a diagonal matrix with weights (expenditure shares, i.e.w_{ij} s or quantities i.e. $q_{ij}^{*} = q_{ij} \Big/ \sum_{m=1}^{M} q_{im}$) in its diagonal. We can incorporate the weights matrix $\mathbf{W}$ in the MOM procedure in a straightforward manner by using weighted moment conditions

$$\mathbf{R}^{'}\mathbf{W}\mathbf{r} = \mathbf{0} \tag{12.15}$$

Obtaining the optimal $\mathbf{R}$ from (12.13) and solving the weighted moments (12.15) leads to the following generalized system[28]

$$\begin{cases} \left\{ PPP_J = \left\{ \sum_{n=1}^{N} w_{nj} \left(\frac{p_{im}}{\widehat{PPP_m}} \right)^{\rho} \right\}^{1/\rho} \right\} & j = 1, \ldots, M-1 \\ \left\{ \hat{P}_i = \left\{ \sum_{m=1}^{M} w_{nj}^{*} \left(\frac{p_{im}}{\widehat{PPP_m}} \right)^{\rho} \right\}^{1/\rho} \right\} & j = 1, \ldots, M-1 \end{cases} \tag{12.16}$$

where following special cases are of interests (see Table 12.1 for more details).

1. *Weighted CPD—Rao system* (12.2) can be obtained by letting $\rho \to 0$ or using logarithmic specification for the CPD model with expenditure share weights.
2. *Arithmetic System:* the weighted arithmetic system (12.4) can be obtained by letting $\rho = 1$ and an expenditure share weight matrix $\mathbf{W}$ and an optimal choiceof $\mathbf{R}$ ala (12.13).

[28] For a discussion of this system and its properties see Hajargasht and Rao (2019).

3. *Harmonic system—IDB system*: the IDB system of Eqs. (12.3) is obtained by letting $\rho = -1$ along with the use of expenditure share weights in matrix **W** and an optimal choice of **R**.
4. *Geary-Khamis system*: the GK system (12.1) can be obtained by letting $\rho = 1$ along with a nonoptimal choice of **R** (see Table 12.1 or Rao & Hajargasht, 2016 for more details) and quantity shares in the weight matrix instead of expenditure shares.

Under homoscedastic assumption $Var(\varepsilon_{ij}) = \sigma^2$, variance of any of the weighted PPPs can be obtained using $Var\left(\left[\widehat{\mathbf{PPP}}, \hat{\mathbf{P}}\right]\right) = \widehat{\sigma^2}\left[\hat{\mathbf{R}}'\mathbf{W}\hat{\mathbf{D}}\right]^{-1}\hat{\mathbf{R}}'\mathbf{WW}\hat{\mathbf{R}}\left[\hat{\mathbf{D}}'\mathbf{W}\hat{\mathbf{R}}\right]^{-1}$ with $\widehat{\sigma^2} = \boldsymbol{r}'\boldsymbol{r}/NM$. Estimation of variances under more general assumptions on the covariance of error terms can be found in Hajargasht and Rao (2021).

Multilateral Comparisons at Basic Heading Level

In this section, we focus on computation of PPPs at the basic heading level and their reliability measures. A distinguishing feature of aggregation at the basic heading level is that expenditure or quantity weights are not available. This section is based on Hajargasht et al. (2019); other relevant references are Ferrari et al. (1996), Ferrari and Riani (1998), Rao (2004), Rao (2013b) and Weinand (2021). We first consider the case where all items are priced in all countries and then we study the more realistic case where not all items are priced in all countries.

Complete Price Matrix: when all items are priced in all countries, it can be shown that the optimal choice for the matrix **R** is

$$\mathrm{R}' = Diag\begin{pmatrix} -\rho/\mathrm{P} \\ -\rho/\mathrm{PPP} \end{pmatrix}\begin{bmatrix} \mathrm{i}_M' \otimes \mathrm{i}_N \\ \mathrm{i}_M \otimes \mathrm{i}_N' \end{bmatrix} \tag{12.17}$$

and

$$R' = Diag\begin{pmatrix} 1/\mathrm{P} \\ 1/\mathrm{PPP} \end{pmatrix}\begin{bmatrix} \mathbf{i}_M' \otimes \mathbf{I}_N \\ \mathbf{I}_M \otimes \mathbf{i}_N' \end{bmatrix} \quad if\ \rho = 0 \tag{12.18}$$

where $\otimes$ denotes Kronecker product and $[.]_{-1}$ represents the matrix [.] without its last row, and $\mathbf{I}_M$ is the identity matrix of size M. Using (12.17) and (12.18) along with (12.11), it can be shown that

$$\begin{cases} \widehat{PPP}_j = \left\{ \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} \left(\dfrac{p_{nj}}{\hat{P}_i} \right)^{\rho} \right\}^{1/\rho} & j = 1, \ldots, M-1 \\ \hat{P}_i = \left\{ \dfrac{1}{M} \displaystyle\sum_{m=1}^{M} \left(\dfrac{p_{im}}{\widehat{PPP}_m} \right)^{\rho} \right\}^{1/\rho} & i = 1, \ldots, N \end{cases} \tag{12.19}$$

where PPP_j s and P_i s can be obtained by solving the above system of equations.[29] ... It can be seen that $\rho = 1$ leads to the arithmetic index; $\rho = 0$ to the Jevons geometric index; and $\rho = -1$ to the harmonic index. To obtain standard errors for PPPs under homoscedasticity assumption on the error term, one can use (12.14) along with the rules for inverse of partitioned matrices to derive for any $\rho \neq 0$

$$Var\left[\hat{\mathbf{P}}, \widehat{\mathbf{PPP}}\right] = \sigma^2 Diag\begin{pmatrix} \hat{\mathbf{P}} / \rho \\ \widehat{\mathbf{PPP}} / \rho \end{pmatrix} \begin{pmatrix} \frac{\mathbf{I}_N}{M} + \frac{(M-1)\mathbf{J}_{N\times N}}{MN} & -\frac{\mathbf{J}_{N\times M-1}}{N} \\ -\frac{\mathbf{J}_{M-1\times N}}{N} & \frac{\mathbf{I}_{M-1}+\mathbf{J}_{M-1\times M-1}}{N} \end{pmatrix} Diag\begin{pmatrix} \hat{\mathbf{P}} / \rho \\ \widehat{\mathbf{PPP}} / \rho \end{pmatrix} \tag{12.20}$$

and for $\rho = 0$

$$Var\left[\hat{\mathbf{P}}, \widehat{\mathbf{PPP}}\right] = \sigma^2 Diag\begin{pmatrix} \hat{\mathbf{P}} \\ \widehat{\mathbf{PPP}} \end{pmatrix} \begin{pmatrix} \frac{\mathbf{I}_N}{M} + \frac{(M-1)\mathbf{J}_{NxN}}{MN} & -\frac{\mathbf{J}_{N\,x\,M-1}}{N} \\ -\frac{\mathbf{J}_{M-1\,x\,N}}{N} & \frac{\mathbf{I}_{M-1}+\mathbf{J}_{M-1\,x\,M-1}}{N} \end{pmatrix} Diag\begin{pmatrix} \hat{\mathbf{P}} \\ \widehat{\mathbf{PPP}} \end{pmatrix} \tag{12.21}$$

where $\mathbf{J}_{N\times M}$ is an N *by* M matrix with all elements equal to one. Note that based on (12.20) and (12.21) with a complete price matrix we have:

$$Var\left(\widehat{\ln \mathrm{PPP}_j}\right) = \frac{2\sigma^2}{N} \tag{12.22}$$

[29] Hajargasht and Rao (2019) discuss the necessary and sufficient conditions for the existence and uniqueness of solutions to the system in (12.19).

Two features of variance of log of PPPs based on the CPD model and with a complete price matrix are notable. First, variances are the same for all basic heading log-PPPs. Second, variances do not depend on the number of countries. As we will see later, these findings do not hold when the price matrix is incomplete.

If estimation of σ^2 is of interest it can be obtained using $\hat{\sigma}^2 = \frac{\hat{r}'\hat{r}}{MN}$. Note that the formula for variance is the same for arithmetic, geometric and harmonic indexes but not for other values of ρ. In what follows we suppress ρ from the equations for convenience but similar results hold when ρ is included.

Reliability measures with Incomplete Price Matrix: a more realistic situation is where not all items are priced in all countries, and therefore, the price matrix is incomplete. To derive standard errors in this case, define a diagonal matrix of dimension $MN \times MN$ denoted by $\mathbf{W} = Diag(d_{ij})$ with d_{ij} equal to 1 if i-th commodity is priced in j-th country, and equal to 0 otherwise. After some algebraic manipulation, it can be shown that

$$Var(\hat{\mathbf{P}}, \widehat{\mathbf{PPP}}) = \sigma^2 Diag\left(\begin{matrix}\hat{\mathbf{P}} \\ \widehat{\mathbf{PPP}}\end{matrix}\right)\left(\mathbf{R}'\,\mathbf{W}\,\mathbf{R}\right)^{-1} Diag\left(\begin{matrix}\hat{\mathbf{P}} \\ \widehat{\mathbf{PPP}}\end{matrix}\right) \quad (12.23)$$

$$\text{with } \mathbf{R}'\mathbf{W}\mathbf{R} = \left[\begin{array}{cccc|cccc} M_1 & 0 & \cdots & 0 & d_{1,1} & . & . & d_{1,M-1} \\ 0 & & & \vdots & . & & & . \\ \vdots & & & 0 & . & & & . \\ 0 & \cdots & 0 & M_N & d_{N,1} & . & . & d_{N,M-1} \\ \hline d_{1,1} & . & . & d_{N,1} & N_1 & 0 & \cdots & 0 \\ . & & & . & 0 & & & \vdots \\ . & & & . & \vdots & & & 0 \\ d_{1,M-1} & . & . & d_{N,M-1} & 0 & \cdots & 0 & N_{M-1} \end{array}\right] \quad (12.24)$$

where $M_i = \sum_{m=1}^{M} d_{im}$ and $N_j = \sum_{n=1}^{N} d_{nj}$. Using the formula for inverse of partitioned matrices it can be shown that

$$\left(\mathbf{R}'\,\mathbf{W}\mathbf{R}\right)^{-1} = \begin{bmatrix} \mathbf{D}_1^{-1} & -\mathbf{A}\mathbf{D}_2^{-1} \\ -\mathbf{D}_2^{-1}\mathbf{A}' & \mathbf{D}_2^{-1} \end{bmatrix}$$

where $\mathrm{D}_1 = Diag(M_i) - \left\{ \sum_{m=1}^{M-1} \frac{d_{i,m} d_{h,m}}{N_m} \right\}_{i,h}$, $\mathrm{D}_2 = Diag(N_j) - \left\{ \sum_{n=1}^{N} \frac{d_{n,j} d_{n,l}}{M_n} \right\}_{j,l}$, $A = \left\{ \frac{d_{i,j}}{M_i} \right\}_{i,j}$ with $i, h = 1, ..., N$ and $j, l = 1, ..., M-1$.

The existence of terms such as $d_{i,m} d_{h,m}$ and M_n suggests that in this case, standard errors of PPPs depend on the degree of overlap of items priced in different countries. Our purpose here is to study this in more detail. First, consider the case where there are only two countries. It is easy to see that in this case

$$Var(\widehat{\ln PPP_1}) = \sigma^2 \mathbf{D}_2^{-1} = \frac{\sigma^2}{\left(N_1 - \sum_{n=1}^{N} \frac{d_{n,1}}{M_n} \right)} \tag{12.25}$$

where N_1 is the number of commodities priced in country one. Note that in this case M_n is either equals 1 or 2. In the two-country case, we have the following special cases:

1. All commodities are priced in both countries:
   $N_1 = N$ and $\sum_{n=1}^{N} \frac{d_{n,1}}{M_n} = \frac{N}{2} \Rightarrow Var(\widehat{\ln PPP_1}) = \frac{2\sigma^2}{N}$
2. No overlap in the commodities priced:

$$\sum_{n=1}^{N} \frac{d_{n,1}}{M_n} = N_1 \Rightarrow N_1 - \sum_{n=1}^{N} \frac{d_{n,1}}{M_n} = 0 \Rightarrow Var(\widehat{\ln PPP_1}) = \infty$$

3. H commodities are in common between the two countries

$$\sum_{n=1}^{N} \frac{d_{n,1}}{M_n} = N_1 - \frac{H}{2} \Rightarrow N_1 - \sum_{n=1}^{N} \frac{d_{n,1}}{M_n} = \frac{H}{2} \Rightarrow Var(\widehat{\ln PPP_1}) = \frac{2\sigma^2}{H}$$

These findings can be summarized as follows: First, the index has the smallest variance where all items are priced in both countries. Second, items priced in just one country have no effect on the variance. Third, when there is no item in common between the two countries, the variance becomes infinite.

Now consider the general case with M countries: In this case, variances depend on the degree of item overlap between countries but in a more complicated way. Note that we can rewrite $\mathbf{D}_2^{-1}$ in the following equivalent form

$$\mathbf{D}_2^{-1} = \begin{bmatrix} \sum_{i=1}^{M-1}\frac{(M-i)N_{1;M-i}}{M-i+1} & -\sum_{i=1}^{M-1}\frac{N_{1,2;M-i-1}}{M-i+1} & \cdots & -\sum_{i=1}^{M-1}\frac{N_{1,M-1;M-i-1}}{M-i+1} \\ -\sum_{i=1}^{M-1}\frac{N_{2,1;M-i-1}}{M-i+1} & \ddots & & \vdots \\ \vdots & & & -\sum_{i=1}^{M-1}\frac{N_{M-2,M-1;M-i-1}}{M-i+1} \\ -\sum_{i=1}^{M-1}\frac{N_{M-1,1;M-i-1}}{M-i+1} & \cdots & -\sum_{i=1}^{M-1}\frac{N_{M-1,M-2;M-i-1}}{M-i+1} & \sum_{i=1}^{M}\frac{M-i}{M-i+1}N_{M-1;M-i} \end{bmatrix}^{-1} \quad (12.26)$$

where $N_{h\,;\,k}$ denotes the number of items that country h has in common with just k other countries and $N_{h\,,\,l\,;k}$ is the number of commodities that country h and l have in common with just k other countries. As it can be seen, matrix $\mathbf{D}_2^{-1}$ depends on $N_{h\,;\,k}$ and $N_{h\,,\,l\,;k}$ which are measures of item overlap between countries. There is no closed form for $\mathbf{D}_2^{-1}$ in general but Hajargasht et al. (2019) consider a number of interesting special cases[30] for which they derive closed form solutions for variances of PPPs. They demonstrate that the reliability of basic heading PPPs depends on price dissimilarities and the degree of item overlap between countries. They also find that a multilateral CPD index always improves precision of PPPs compared to a bilateral comparison although this conclusion may not hold when one goes beyond the homoscedasticity assumption.

The Stochastic Approach to GEKS

GEKS is the main aggregation method at above basic heading level in ICP (Rao, 2013a; World Bank, 2020). In this section, we further discuss the GEKS index and then show how its reliability measure can be estimated. This section is based on Hajargasht (2021) and Hajargasht and Rao

[30] See Fig. 12.1 in Hajargasht, Rao and Valadkhani (2019).

(2021). Cuthbert (2003) and Deaton (2012) are two other references considering computation of reliability measures for GEKS.

GEKS starts from the premise that the best comparison between a pair of countries is an ideal bilateral index and then it tries to obtain a transitive multilateral index that is the "closest" to this bilateral index. Usually, the Fisher index is used since it is considered to be the ideal bilateral index.[31] The Fisher index F_{jk} is defined with $F_{jk} = \sqrt{L_{jk} P_{jk}} = \sqrt{\frac{\sum_{n=1}^{N} p_{nj} q_{nk}}{\sum_{n=1}^{N} p_{nk} q_{nk}} \frac{\sum_{n=1}^{N} p_{nj} q_{nj}}{\sum_{n=1}^{N} p_{nk} q_{nj}}}$ where L_{jk} and P_{jk} denote Laspeyres and Paasche indexes between country j and k. The GEKS index between country j and k is then derived by minimizing the logarithmic least squares

$$Min_{PPP_{jk}} \sum_{j=1}^{M} \sum_{k=1}^{M} \left(\ln PPP_{jk}^{GEKS} - \ln F_{jk} \right)^2 \tag{12.27}$$

subject to the transitivity condition that $PPP_{jk}^{GEKS} = PPP_{jl}^{GEKS} PPP_{lk}^{GEKS}$ for every j, k and l. It can be shown (see, e.g., Rao, 2009) that PPP_{jk}^{GEKS} satisfies transitivity if and only if there exists $\{PPP_1, PPP_2, ...PPP_M\}$ such that $PPP_{jk}^{GEKS} = PPP_j / PPP_k$. Therefore, Eq. (12.27) can be also written as

$$Min_{PPP_j, PPP_k} \sum_{j=1}^{M} \sum_{k=1}^{M} \left(\ln PPP_j - \ln PPP_k - \ln F_{jk} \right)^2 \tag{12.28}$$

For this to have a unique solution, a normalization is needed often of the form $PPP_M = 1$. It is not difficult to see that the solutions, $\ln PPP_j$ and $\ln PPP_k$ are also the ordinary least squares (OLS) estimators from

$$\ln F_{jk} = \ln PPP_j - \ln PPP_k + \varepsilon_{jk} \tag{12.29}$$

[31] While Fisher price indexes are usually used in the GEKS transitivization, the Tornqvist price index has also been used (see Caves et al., 1982). Any other binary index number formula that satisfies time or country reversal test can also be used (see e.g. Rao & Banerjee, 1986). The GEKS index which uses the Tornqvist binary index is often referred to as CCD. See Rao and Selvanathan (1991) for a regression-based stochastic approach to CCD index.

where ε_{jk} is a random error term. Minimizing (12.28) or applying least squares to (12.29) along with the fact that $PPP_{jk}^{GEKS} = \frac{PPP_j}{PPP_k}$, it can be shown that

$$\begin{aligned} \ln PPP_{jk}^{GEKS} &= \frac{1}{M}\sum_{l=1}^{M}[\ln F_{jl} + \ln F_{lk}] \\ \text{or} \quad PPP_{jk}^{GEKS} &= \prod_{l=1}^{M}\left[F_{jl}\cdot F_{lk}\right]^{1/M} \end{aligned} \tag{12.30}$$

One way of obtaining the reliability measure for GEKS is by direct application of the formula for variance of sums to (12.30) and write

$$\begin{aligned} &Var(\widehat{\ln PPP_{jk}^{GEKS}}) \\ &= \frac{1}{M^2}\left\{\begin{array}{l} \sum_{l=1}^{M}\left[\begin{array}{l} Var(\widehat{\ln F_{jl}}) + Var(\widehat{\ln F_{lk}}) \\ +2Cov(\widehat{\ln F_{jl}}, \widehat{\ln F_{lk}}) \end{array}\right] \\ +2\sum_{l=1}^{M}\sum_{h\neq l}^{M}\left[Cov(\widehat{\ln F_{jl}}, \widehat{\ln F_{jh}}) + Cov(\widehat{\ln F_{lk}}, \widehat{\ln F_{hk}})\right] \end{array}\right\} \end{aligned} \tag{12.31}$$

or in matrix form $\widehat{\ln \mathrm{PPP}_{jk}^{GEKS}} = \frac{1}{M}\mathrm{i}_{2M}' \begin{bmatrix} \widehat{\ln \mathrm{F}_j} \\ \widehat{\ln \mathrm{F}_k} \end{bmatrix}$ and

$$\begin{aligned} Var(\widehat{\ln \mathrm{PPP}_{jk}^{GEKS}}) &= Var\left\{\frac{1}{M}\mathrm{i}_{2M}'\begin{bmatrix} \widehat{\ln \mathrm{F}_{j\_}} \\ \widehat{\ln \mathrm{F}_{\_k}} \end{bmatrix}\right\} \\ &= \frac{1}{M^2}\mathrm{i}_{2M}' Var\left\{\begin{bmatrix} \widehat{\ln \mathrm{F}_{j\_}} \\ \widehat{\ln \mathrm{F}_{\_k}} \end{bmatrix}\right\}\mathrm{i}_{2M} \end{aligned} \tag{12.32}$$

where $\mathbf{i} = (1, \cdots, 1)', F_{j\_} = (F_{j1}, \cdots, F_{jM})'$ and $F_{\_k} = (F_{1k}, \cdots\cdots, F_{Mk})'$. Formula (12.31) or (12.32) requires variances and covariances of the Fisher index. Hajargasht (2021) obtained the following formulas for variances and covariances of the logarithm of the Fisher index.

$$Var\left(\widehat{\ln F_{jk}}\right) = \frac{1}{4}\sum_{n=1}^{N}\left\{w_{nk}\left(\frac{p_{nj}}{p_{nk}L_{jk}} - 1\right) - w_{nj}\left(\frac{p_{nk}P_{jk}}{p_{nj}} - 1\right)\right\}^2 \tag{12.33}$$

$$Cov\left(\widehat{\ln F_{jk}}, \widehat{\ln F_{lm}}\right) = \frac{1}{4}\sum_{n=1}^{N}\left\{w_{nk}\left(\frac{p_{nj}}{p_{nk}L_{jk}} - 1\right) - w_{nj}\left(\frac{p_{nk}P_{jk}}{p_{nj}} - 1\right)\right\}\left\{w_{nl}\left(\frac{p_{nl}}{p_{nm}L_{lm}} - 1\right) - w_{nm}\left(\frac{p_{nm}P_{lm}}{p_{nl}} - 1\right)\right\} \tag{12.34}$$

where L_{jk} and P_{jk} denote Laspeyres and Paasche indexes between country j and k. Substituting expressions (12.33) and (12.34) into (12.31) provides the variance for the GEKS index. As it can be seen, computing reliability measures for GEKS is straightforward when formulas for variance and covariances of the Fisher index are available.

Conclusion

This paper provided an overview of the stochastic approach to index numbers for international price comparisons. It drew heavily from Rao (2004, 2005), Rao and Hajargasht (2016), Hajargasht, Rao and Valadkhani (2019) and Hajargasht and Rao (2010, 2021) which is a testament to Prasada Rao's significant contribution to the development of stochastic approach to multilateral price comparisons. The review focused mainly on providing an exposition of the new stochastic approach to multilateral price comparisons. For applications of these methods to real data and empirical comparison of the indexes, the reader is referred to the papers cited above.

The earlier version of the stochastic approach to index numbers had ad-hoc features and therefore was criticized. But as this survey demonstrates, there has been a shift in the specification of the regression model that underpins stochastic approach to price index numbers and significant progress has been made in recent years. Thanks to the use of more modern statistical tools, it is now possible to specify sound stochastic models that can generate unweighted elementary systems such as the Jevons, Dutot, arithmetic and harmonic indexes as well as weighted indexes such as Rao, IDB, GK and GEKS methods for multilateral comparisons. This chapter has also provided an overview of the methods for computing reliability measures or standard errors for all of the indexes used in the ICP when appropriate covariance structures for the error term are specified.

Despite the progress, outstanding problems remain to be studied. We can now compute reliability measures for PPP indexes but how should these measure be best interpreted? Accounting for correlation among prices deemed to be important since it could impact the estimated PPPs and their reliability measures (see, e.g., Aten, 1996; Rao, 2004, 2009). Another topic worthy of further investigation is the use of resampling methods such as the bootstrap in computing reliability measures. While the ultimate aim in international comparisons is comparing quantities (e.g. real incomes), an appropriate stochastic approach to quantity indexes is lacking. Hajargasht (2021) and Hajargasht and Rao (2021) have tried to address some of these issues. Extension of the stochastic approach to subnational PPPs (e.g. Laureti & Rao, 2018; Majumder & Ray, 2020), application of multilateral indexes to scanner data (e.g. Diewert and Fox, 2020; Laureti et al., 2017), the use of bilateral reliability measures in constructing more reliable multilateral indexes (e.g. Hajargasht et al., 2018; Hill & Timmer, 2006), and accounting for differences in the quality of products and services across regions are some of the other important area for further research.

References

Aten, B. H. (1996). Evidence of spatial autocorrelation in international prices. *Review of Income and Wealth, 42*(2), 149–163.

Aldrich, J. (1992). Probability and depreciation: A history of the stochastic approach to index numbers. *History of Political Economy, 24*, 657–687.

Balk, B. M. (1980). A method for constructing price indices for seasonal commodities. *Journal of the Royal Statistical Society A, 143*, 68–75.

Balk, B. M. (1996). A comparison of ten methods for multilateral international price and volume comparisons. *Journal of Official Statistics, 12*, 199–222.

Balk, B. M. (2008). *Price and quantity index numbers: Models for measuring aggregate change and difference*. Cambridge University Press.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal, 92*(365), 73–86.

Clements, K. W., & Izan, H. Y. (1981). A note on estimating divisia index numbers. *International Economic Review, 22*, 745–747.

Clements, K. W., & Izan, H. Y. (1987). The measurement of inflation: A stochastic approach. *Journal of Business & Economic Statistics, 5*(3), 339–350.

Clements, K. W., Izan, H. Y., & Selvanathan, E. A. (2006). Stochastic index numbers: A review. *International Statistical Reviews, 74*(2), 235–270.

Cuthbert, J. R. (2003). On the variance-covariance structure of the log Fisher index, and implications for aggregation techniques. *Review of Income and Wealth, 49*(1), 69–88.

Deaton A. (2012). Calibrating measurement uncertainty in purchasing power parity exchange rates. *Paper Presented to the World Bank's ICP Technical Advisory Group.*

Deaton, A., & Heston, A. (2010). Understanding PPPs and PPP based national accounts. *American Economics Journal: Macroeconomics., 2*(4), 1–35.

Diewert, W. E. (1995). *On the stochastic approach to index numbers* (Discussion Paper No. 95–31). Department of Economics, University of British Columbia.

Diewert, W. E. (1999). Axiomatic and economic approaches to international comparisons. In A. Heston & R. E. Lipsey (Eds.), *International and interarea comparisons of income* (pp. 13–87). Output and Prices, NBER, University of Chicago Press.

Diewert, W. E. (2004). On the stochastic approach to linking the regions in the ICP (Department of Economics, Discussion Paper 04–16). University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1.

Diewert, W. E. (2005). Weighted country product dummy variable regressions and index number formulae. *The Review of Income and Wealth, 51*(4), 561–571.

Diewert, W. E. (2013). Methods of aggregation above the basic heading level within regions. In W. Bank (Ed.), *Measuring the real size of the world economy* (pp. 121–167). World Bank (Chapter 5).

Diewert, W. E. (2020). The stochastic approach to index numbers. https://econ2017.sites.olt.ubc.ca/files/2020/04/pdf_paper_erwin-diewert_DP2004StochasticApproachestoIndexNumberTheory.pdf

Diewert, W. E., & Fox, K. J. (2020). Substitution bias in multilateral methods for CPI construction. *Journal of Business and Economic Statistics*. https://doi.org/10.1080/07350015.2020.1816176

Dikhanov, Y. (1994, August 21–27). Sensitivity of PPP-based income estimates to the choice of aggregation procedures. The World Bank, Washington D.C., June 10, *paper presented at* 23rd *General Conference of the International Association for Research in Income and Wealth*, St. Andrews, Canada.

Dikhanov, Y. (2020). *The Gerschenkron effect in international comparisons, 2011 and 2017*. 2–02-RA-Item-05-The-Gerschenkron-effect-in-ICP-2011-and-ICP-2017-Dikhanov-2.pdf

Eltetö, O., & Köves, P. (1964). On a problem of index number computation relating to international comparison. *Statisztikai Szemle, 42*, 507–518.
Ferrari, G., G. Gozzi, G., & Riani, M. (1996). Comparing CPD and GEKS approaches at the basic headings level. In Eurostat (Ed.), *CPI & PPP: Improving the quality of price indices* (pp. 323–337).
Ferrari, G., & Riani, M. (1998). On purchasing power parities calculation at the basic heading level. *Statistica, 58*(1), 91–108.
Geary, R. C. (1958). A note on the comparison of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society, Series A, 121*(1), 97–99.
Gini, C. (1931). On the circular test of index numbers. *International Review of Statistics, 9*(2), 3–25.
Gorajek, A. (2018). Econometric perspectives on economic measurement. *Econometric BA Research Discussion Papers rdp2018–08, Reserve Bank of Australia*.
Hajargasht, G. (2021). Variance of ideal price and quantity indexes. *mimeographed*.
Hajargasht G., Hill, R. J., Rao, D. S. P., & Shankar, S. (2018). *Spatial chaining in international comparisons of prices and real incomes (No. 2018–03).* University of Graz, Department of Economics.
Hajargasht, G., & Rao, D. S. P. (2010). Stochastic approach to index numbers for multilateral price comparisons and their standard errors. *Review of Income and Wealth, Series, 56*(2010), S32–S58.
Hajargasht, G., & Rao, D. S. P. (2019). Multilateral index number systems for international price comparisons: Properties, existence and uniqueness. *Journal of Mathematical Economics, 83*, 36–47.
Hajargasht, G., & Rao, D. S. P (2021). Constructing reliability measures for purchasing power parities. *Paper presented in ICP TAG Meeting*.
Hajargasht, G., Rao, D. S. P., & Valadkhani, A. (2019). Reliability of basic heading PPPs. *Economics Letters, 180*, 102–107.
Hill, R. J. (1997). A taxonomy of multilateral methods for making international comparisons of prices and quantities. *Review of Income and Wealth, 43*(1), 49–69.
Hill, R. J., & Syed, I. A. (2015). Improving international comparisons of prices at basic heading level: An application to the Asia-Pacific region. *Review of Income and Wealth, 61*(3), 515–539.
Hill, R. J., & Timmer, M. P. (2006). Standard errors as weights in multilateral price indexes. *Journal of Business and Economic Statistics, 24*(3), 366–377.
Iklé, D. M. (1972). A new approach to the index number problem. *The Quarterly Journal of Economics, 86*(2), 188–211.
Keynes, J. M. (1930). *A teatise on money,volume I: The pure theory of money.* Macmillan.

Khamis, H. (1972). A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society, Series A, 135*(1), 61–85.

Kravis, I. B., Heston, A., & Summers, R. (1982). *World product and income: International comparisons of real gross product*. Johns Hopkins University Press.

Kravis, I. B., Kenessey, Z., Heston, A., & Summers, R. (1975). *A system of international comparisons of gross product and purchasing power*. The World Bank.

Laureti, T., Ferrante, C., & Dramis, D. (2017). Using scanner and CPI data to estimate Italian Sub-national PPPs. In *Statistics and data science: New challenges, new generations*.

Laureti, T., & Rao, D. S. P. (2018). Measuring spatial price level differences within a country: Current status and future developments. *Studies of Applied Economics, 36*(1), 119–148.

Majumder, A., & Ray, R. (2020). National and subnational purchasing power parity: A Review. *Decision, 47*(2), 103–124.

Montero, J. M., Laureti, T., Mínguez, R., & Fernández-Avilés, G. (2020). A stochastic model with penalized coefficients for spatial price comparisons: An application to regional price indexes in Italy. *Review of Income and Wealth, 66*(3), 512–533.

Rao, D. S. P. (1990). A system of log-change index numbers for multilateral comparisons. In Salazar-Carrillo and Rao (Eds.), *Comparisons of prices and real products in Latin America*. Contributions to Economic Analysis Series, Amsterdam: North-Holland Publishing Company.

Rao, D. S. P. (1995). *On the equivalence of the Generalized Country-product-dummy (CPD) method and the rao-system for multilateral comparisons* (Working Paper No. 5). Centre for International Comparisons, University of Pennsylvania, Philadelphia.

Rao, D. S. P (2004). The country-product-dummy method: A stochastic approach to the computation of purchasing power parities in the ICP. In *SSHRC Conference on Index Numbers and Productivity Measurement* (Vol. 30).

Rao, D. S. P. (2005). On the equivalence of weighted country–product–dummy (CPD) method and the rao-system for multilateral price comparisons. *Review of Income and Wealth, 51*, 571–580.

Rao, D. S. P. (2009). Generalised Elteto-koves-Szulc (EKS) and country-product-dummy (CPD) methods for international comparisons. In P. Power (Ed.), *Prasada Rao* (pp. 86–120). Recent Advances In Methods And Applications, Edward Elgar Publishing Company.

Rao, D. S. P. (2013a). The framework for the international comparison program (ICP). Chapter 1 in World Bank (ed) *Measuring the real size of the world economy* (pp. 13–45). World Bank.

Rao, D. S. P. (2013b). Computation of basic heading purchasing power parities (PPPs) for comparisons within and between regions. Chapter 4 in World Bank (Ed.), *Measuring the real size of the world economy* (pp. 93–119). World Bank.

Rao, D. S. P., & Banerjee, K. S. (1986). A multilateral system of index numbers based on factorial approach. *Statistische Hefte, 27*, 297–312.

Rao, D. S. P., & Hajargasht, G. (2016). Stochastic approach to computation of purchasing power parities in the international comparison program (ICP). *Journal of Econometrics, 191*(2), 414–425.

Rao, D. S. P., & Selvanathan, E. A. (1991). A log-change index number formula for multilateral comparisons. *Economics Letters, 35*(297–300), 1991.

Rao, D. S. P., & Selvanathan, E. A. (1992). Computations of standard errors for Geary-Khamis parities and international prices. *Journal of Business and Economics Statistics, 10*, 109–115.

Selvanathan, E. A., & Rao, D. S. P. (1994). *Index numbers: A stochastic approach.* University of Michigan Press.

Summers, R. (1973). International price comparisons using incomplete data. *Review of Income and Wealth, 19*, 1–6.

Szulc, B. (1964). Indices for multiregional comparisons. *Przeglad Statystyczny, 3*, 239–254.

Theil, H. (1967). *Economics and information theory*. North-Holland Publishing.

Walsh, C. M. (1924). Professor Edgeworth's views on index-numbers. *Quarterly Journal of Economics, 38*, 500–519.

Weinand, S. (2021), Measuring spatial price differentials at the basic heading level: A comparison of Stochastic index number methods. *AStA Advances in Statistical Analysis*, 1–27.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

World Bank. (2013). *Measuring the real size of the world economy: The framework, methodology, and results of the international comparison program—ICP*. World Bank.

World Bank. (2020). *Purchasing power parities and the size of world economies: Results from the 2017 international comparison program.* The World Bank.

CHAPTER 13

Inconsistencies in Cross-Country Price Comparisons over Time: Patterns and Facts

Robert Inklaar, Ryan Marapin, Pieter Woltjer, and Marcel Timmer

R. Inklaar (✉) · R. Marapin · P. Woltjer · M. Timmer
Groningen Growth and Development Centre, University of Groningen, Groningen, The Netherlands
e-mail: r.c.inklaar@rug.nl

R. Marapin
e-mail: ryan.marapin@rug.nl

P. Woltjer
e-mail: p.j.woltjer@rug.nl

M. Timmer
e-mail: m.p.timmer@rug.nl

M. Timmer
Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands

D. Chotikapanich et al. (eds.), *Advances in Economic Measurement*, https://doi.org/10.1007/978-981-19-2023-3_13

Introduction

Statistical programmes for measuring purchasing power parities (PPPs), such as the International Comparison Program (ICP), are designed to give an accurate snapshot of comparative price levels across countries at a point in time. That snapshot, in turn, informs judgements about the comparative size of economies, the extent of absolute poverty and the international income distribution; see World Bank (2020) for the most recent global price comparison for 2017 and Deaton and Schreyer (2021) for a broader perspective on these price comparisons. Global price comparisons by ICP started in the 1970s with few participating countries and have since expanded to cover more than 170 countries in the most recent global comparisons.

This series of cross-country price level comparisons can be set against measures of inflation, i.e. price changes within countries, over time. Since PPPs and inflation are both measures based on prices of products in different countries and at different points in time, we would expect a degree of consistency between these two measures. More precisely, we might expect that the change over time in the PPP between two countries is close to the relative rate of inflation.[1] Yet in practice, the change in PPPs over time can be very different from patterns of relative inflation between countries, see e.g. Deaton (2010) or Inklaar and Rao (2017).

This points to deeper problems in the comparability of products and prices that are compared in national and international surveys, for instance, due to methodological differences in price measurement (Deaton & Aten, 2017; Hill, 2004; McCarthy, 2013). The inconsistency between PPPs and inflation might have serious consequences for economic research since the choice for a specific snapshot, or PPP benchmark year, can impact estimates of international income inequality or affect the results of entire studies.[2] And since international databases, such as the World Development Indicators and the Penn World Table, rely on (timelier) inflation data to extrapolate PPPs to the most recent years, the release of new ICP data can lead to large revisions to relative price and income levels.

[1] Inconsistencies due to the use of different weights are to be expected, but these would be relatively small, see e.g. Inklaar and Timmer (2013), McCarthy (2013) and the exposition in the next section of this paper based on Deaton and Aten (2017).

[2] See Johnson et al. (2013) and Ciccone and Jarociński (2010).

The first substantive discussion of this issue goes back to Krijnse Locker and Faerber (1984). The Ryten report (United Nations Statistical Commission, 1999) also flags 'incoherence' (inconsistency in our terminology) as a matter of serious concern for the quality and credibility of the ICP. Summers and Heston (1991) discussed a so-called consistentisation approach for the Penn World Table (PWT) to reconcile time-series of comparative price and income levels compiled based on different PPP benchmarks. More recent examples of methods to reconcile PPP benchmarks and inflation are the approaches of Rao et al. (2010) and Hill and Melser (2015). Yet this emphasis on reconciliation skips a step, in our view, because we do not yet have a very good understanding of the scope and features of the underlying problem of inconsistency between PPP changes and relative inflation.

As a step towards remedying this, we document a new set of stylised facts about these inconsistencies in this paper. The aim is to provide a systematic perspective on where and when inconsistencies are larger. This can help inform more formal modelling of the type done by Rao et al. (2010) and provide a view on the uncertainty surrounding PPP estimates in other settings as well, for instance, to use these as weights in econometric analysis. By identifying where inconsistencies are largest, we can also help point to specific measurement challenges with the aim of improving statistical practice. We build on earlier research that focused on comparing specific benchmarks, such as Deaton and Aten (2017) and Inklaar and Rao (2017) who compare the ICP 2005 and ICP 2011 results, but we propose a more general framework for analysing inconsistency. This framework is motivated by some of the specific difficulties for comparing prices across countries that Deaton and Heston (2010) discuss.

To frame the issue more clearly, consider the stylised example illustrated in Fig. 13.1. For comparing prices between countries A and B, there are three PPP benchmarks, in years 1, 2 and 3 plotted in the top-left panel. We also observe inflation in both countries and can thus plot inflation in country A relative to country B in the bottom-left panel. In the right panel, these two sets of information are combined. Starting from each benchmark, an estimate can be made of PPPs in subsequent years using relative inflation. Applying these estimates to each of the three benchmarks leads to the parallel lines shown in the right panel, with the blue line starting from the benchmark in year 1, the red line in year 2 and the green line in year 3.

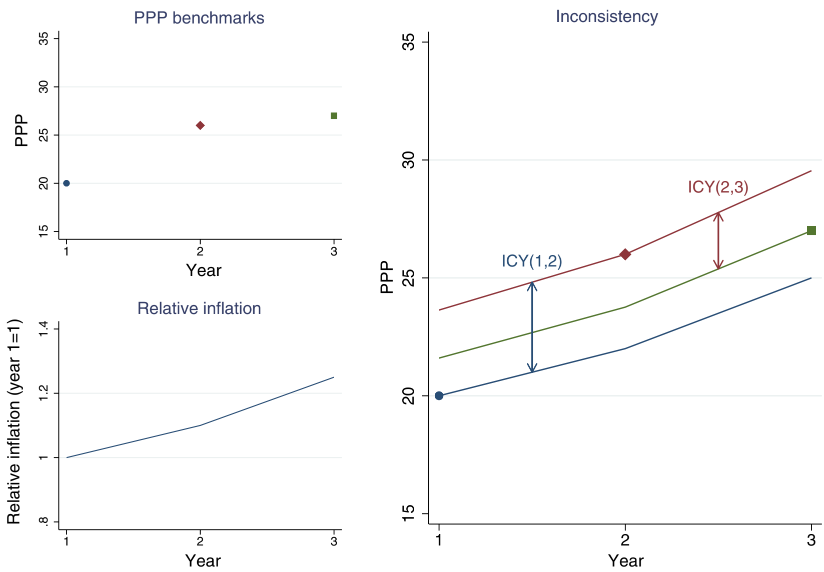

Fig. 13.1 A stylised example of inconsistency between PPP benchmarks and relative inflation

The distance between two of the lines is our measure of inconsistency, with ICY (1, 2) referring to the inconsistency between the PPP benchmark in year 1 and year 2 and ICY (2, 3) the inconsistency between benchmarks in year 2 and 3.[3] In the next section, we will provide a more formal framework for measuring inconsistency and comparing the degree of inconsistency in a setting of multiple countries, benchmarks and years. In general, inconsistencies can be due to differences in the expenditure shares across countries—with inflation based only on national shares and PPPs on shares for multiple countries—due to differences in the measurement method for inflation versus PPPs or due to differences in product sampling in national and international price surveys (Deaton & Aten, 2017; Hill, 2004; McCarthy, 2013).

[3] This general framework is also at the basis of the work of Rao et al. (2010), whose final real income series is based on a weighted average of the three lines shown in the right panel of Fig. 13.1.

The questions we ask of the data are primarily inspired by Deaton and Heston (2010), who highlight the difficulty of comparing prices across disparate countries:

1. *Are comparisons between more recent global PPP benchmarks more consistent?* Given the more extensive resources devoted to the recent benchmarks, as well as the greater methodological refinement, we would expect that comparing two recent PPP benchmarks will show smaller inconsistencies than comparing two earlier PPP benchmarks. At the same time, the more recent benchmarks are more extensive in scope, covering over 170 countries since 2011 versus 115 in 1996 and even fewer before that. This increase in scope could lead to greater difficulties in comparing like-with-like.
2. *Is there more consistency between more similar countries?* Making a global price comparison requires comparing very disparate countries. This will be more difficult if expenditure patterns across products are very different, because a price comparison needs to bridge that gap in spending patterns. It will also be more difficult to compare prices of identical products, as, for example, the quality of housing or of schooling can be very different. This is an important reason why some argue for building a global income comparison by comparing more similar countries in a stage-wise comparison instead a single multilateral comparison (e.g. Hajargasht et al., 2018).
3. *Is there more consistency for products that are easier to price and compare across countries?* At a high aggregation level, we expect less inconsistency for household consumption than for other major expenditure categories. At a more detailed level, ICP has always emphasised that certain parts of GDP are 'comparison-resistant', such as public services, housing and construction. There is also a sizeable number of expenditure categories where ICP makes no direct price observations, instead relying on PPPs for other categories or higher aggregates. We expect that it is more likely that the evolution of (e.g.) food PPPs matches relative food prices than that the PPP for public services matches the implicit price deflator for this expenditure category.
4. *Would more frequent benchmark comparisons lead to more consistency?* A longer period between comparisons will lead to larger differences in product samples and expenditure patterns, which could lead to greater inconsistency. Statistical practice is moving towards more

frequent comparisons, with annual PPP estimates across Europe and ICP also shifting from a six-year gap to a three-year gap.[4]

5. *Do inconsistencies distort the international income distribution?* Especially in the papers comparing ICP 2005 to ICP 2011, a major concern was that inconsistencies were larger in lower-income countries, implying larger global cross-country income inequality based on ICP 2005 than on ICP 2011. While any inconsistency makes it hard to establish income rankings of countries, if these inconsistencies vary systematically with income levels, even assessing global inequality trends will be difficult.

We use the most recent version of the Penn World Table, 10.0 (Feenstra et al., 2015) to analyse patterns in inconsistency starting with results from the first ICP comparison for 1970. This version is particularly suited for the purpose of this paper because it incorporates the results of the most recent ICP benchmarks, including that of 2017 (World Bank, 2020). If there is a trend over time towards more consistency, being able to compare the ICP 2011 and ICP 2017 benchmarks is important as these have not attracted the type of criticism that the earlier benchmarks have. We supplement PWT data for the 1970–2017 period with information on PPPs and relative inflation for detailed product categories, the basic-heading level of ICP. These more detailed data are used in the construction of the ICP PPPs and cover the period 2011–2017.

We define inconsistency based on Fig. 13.1 as the distance (in log terms) between the parallel lines in the right-hand side panel. Our findings on the questions we formulated above are as follows:

1. More recent ICP benchmarks are less inconsistent, pointing to the importance of improved measurement methods.
2. Price comparisons between countries:
    (a) With more similar expenditure patterns are less inconsistent.
    (b) With more similar income levels are (frequently) less inconsistent.

[4] ICP was due to conduct a global comparison for 2020 before the COVID-19 pandemic hit and that made the requisite data collection much harder. A new comparison is planned for 2021.

3. When comparing inconsistency by household consumption expenditure category, we find that some harder-to-measure product categories such as education and housing have high degrees of inconsistency and some easier-to-measure categories such as food products and clothing are lower. Yet inconsistency is also high for some categories without major measurement challenges, such as furnishing and household equipment.
4. When comparing PPPs across multiple benchmarks, there is no clear upward trend in inconsistency. We would have expected such an upward trend if differences in spending patterns lead to an accumulation of inconsistencies over time. This result could instead indicate that random factors are predominant in driving inconsistency. However, the modest number of benchmark comparisons make firm conclusions on this hard to draw.
5. The only PPP benchmark where inconsistency varied systematically with income was ICP 1980.[5] As a result, the PPPs from ICP 1980 show a higher degree of income inequality than what is implied by extrapolating PPPs from ICP 1975 forward or from ICP 1985 backwards.

Our paper relates to the work by Rao and Hajargasht (2016) on estimating standard errors for PPPs. The approach taken in that paper (and related literature) is to use the variation in prices for individual items around the (weighted) average price level (i.e. the PPP) as indicative of the uncertainty surrounding the PPP. In our analysis, we try to quantify PPP uncertainty by comparing changes in PPPs to relative inflation. We focus on a different aspect of 'mismeasurement' because inconsistency between PPP changes and relative inflation can be due to mismeasured inflation as well as mismeasured PPPs. Index number problems can also drive inconsistencies—PPPs are estimated using expenditure patterns for multiple countries while inflation is only based on domestic expenditure patterns. At the same time, some of the price variation that Rao and Hajargasht (2016) analyse can be traced to systematic cross-country

[5] Rather than using the official ICP 2005 results, we use PPP data from PWT, which incorporates the adjustments proposed by Inklaar and Rao (2017) to correct for methodological differences and biases.

differences in price patterns. Most notably, the Balassa-Samuelson hypothesis predicts lower relative prices for services than for goods in low-income countries compared to high-income countries.

Our analysis of inconsistency over longer periods of time is especially relevant in a historical context. Since the work of Maddison (2001, 2007), reliance on a single global comparison for 1990 has been the dominant approach (see e.g. Bolt et al., 2018). This is despite growing evidence that this modern price comparison is inconsistent with historical price comparisons (e.g. Veenstra, 2015; Woltjer, 2015) or the price-income relationship that can be seen in every international price comparison (Prados de la Escosura, 2000). Understanding the degree of inconsistency especially over longer periods of time can be helpful to make sense of inconsistencies between modern (i.e. post-1950) price comparisons and historical comparisons, such as by Ward and Devereux (2021).

Conceptual Framework

Our goal is to compare the consistency of PPP estimates and relative inflation using price and expenditure information for multiple products and countries. To frame this issue more clearly, let us first consider the price p of an individual product i in a two-country setting, country j relative to country k. The PPP for that product at time t is then defined as:

$$\mathrm{PPP}_{ijk}^{t} = \frac{p_{ij}^{t}}{p_{ik}^{t}} \tag{13.1}$$

Next define the rate of price change over time, π, for that same product between time v and time t in country j (and k):

$$\pi_{ij}^{vt} = \frac{p_{ij}^{t}}{p_{ij}^{v}} \tag{13.2}$$

Given these definitions, the change in PPP between time v and time t must be equal to the relative rate of inflation between the two countries over the same period:

$$\frac{\mathrm{PPP}_{ijk}^{t}}{\mathrm{PPP}_{ijk}^{v}} = \frac{p_{ij}^{t}}{p_{ik}^{t}} / \frac{p_{ij}^{v}}{p_{ik}^{v}} = \frac{\pi_{ij}^{vt}}{\pi_{ik}^{vt}} \tag{13.3}$$

Again, we are focusing here on the price for a single product, say a bag of rice, so Eq. (13.3) must hold.

Comparing the change in PPPs to relative inflation at higher levels of aggregation complicates the equality from Eq. (13.3) for three reasons, namely that:

1. Aggregate PPPs and inflation are an aggregate of individual product prices using expenditure weights,
2. PPPs and inflation are sometimes measured in different ways for the same product (category), and
3. PPPs and inflation are based on different samples of products.

Reasons 2 and 3 will be discussed at greater length in the empirical sections but note already that these reasons can guide expectations on where the inconsistencies are expected to be larger. That in turn may provide the grounds for ranking these measurement challenges in order of importance.

The first reason, related to expenditure weights, is explained well in Deaton and Aten (2017, 251), whose exposition we follow here. When calculating inflation across multiple products, the price changes of individual products should be weighted by their share in the expenditure basket in that country. Assume, for expositional simplicity, that expenditure shares s differ across countries but remain constant over time. Using a Törnqvist index, we can write the cross-country difference in overall inflation rate π^{vt} as:

$$\Delta \log \pi_j^{vt} - \Delta \log \pi_k^{vt} = \sum_i \left(s_{ij} \Delta \log \pi_{ij}^{vt} \right) - \sum_i \left(s_{ik} \Delta \log \pi_{ik}^{vt} \right) \quad (13.4)$$

The Törnqvist PPP at time t can, in turn, be written as:

$$\log \text{PPP}_{jk}^{t} = \sum_i \frac{1}{2} \left(s_{ij} + s_{ik} \right) \log \frac{p_{ij}^t}{p_{ik}^t} \quad (13.5)$$

Combining Eqs. (13.4) and (13.5), we can write the change in PPPs as:

$$\Delta \log \text{PPP}_{jk}^{vt} = \left(\Delta \log \pi_j^{vt} - \Delta \log \pi_k^{vt} \right)$$

$$- \sum_i \frac{1}{2}(s_{ik} - s_{ij})\left(\Delta \log \pi_{ij}^{vt} + \log \pi_{ik}^{vt}\right) \quad (13.6)$$

The first term in brackets is the log approximation to Eq. (13.3),[6] but added to this is the second term, which introduces a systematic difference between the change in PPP over time and relative inflation. This term will be larger when expenditure shares differ more between the countries and when a product has a higher average inflation rate. Note that this discussion refers to inflation, following the consumption-oriented discussion of Deaton and Aten (2017), but this discussion applies equally to other price indexes and deflators, up to the level of the GDP deflator.

As discussed above, the effect of differences in expenditure shares in Eq. (13.6) is only one of the three factors that may be relevant in practice. In general, we define the degree of inconsistency between PPP changes and relative inflation, d, as:

$$d_{jk}^{vt} \equiv \Delta \log \mathrm{PPP}_{jk}^{vt} - \left(\Delta \log \pi_j^{vt} - \Delta \log \pi_k^{vt}\right) \quad (13.7)$$

We express the inconsistency in logarithmic form so that the measure is symmetric between countries and time periods, i.e. $d_{jk}^{\tau t} = -d_{jk}^{t\tau} = -d_{kj}^{\tau t} = d_{kj}^{t\tau}$. In a two-country setting with two periods, $d_{jk}^{\tau t}$ provides a complete description of inconsistency, but with multiple countries and multiple PPP benchmarks, it is useful to define summary measures, as in Inklaar and Rao (2017). Our main summary measure is the root mean squared inconsistency RMSI:

$$\mathrm{RMSI}^{vt} = \left(\frac{1}{C}\sum_j \left(d_{jk}^{vt} - \overline{d}_k^{vt}\right)^2\right)^{\frac{1}{2}} \quad (13.8)$$

Here $\overline{d}_k^{vt} = \frac{1}{C}\sum_j d_{jk}^{vt}$ is the average inconsistency over the set of countries C. This measure is based on the inconsistencies for a given base-country k, but thanks to the symmetry of the inconsistency measure d, the RMSI^{vt} measure is base-country independent.

[6] The correspondence is only exact in continuous time.

We will also consider the slope coefficient from regressing log income level on inconsistencies:

$$d_{jk}^{vt} \equiv \alpha + \beta^{vt} \log y_j + \varepsilon_j \tag{13.9}$$

Here, again, we choose a base-country k but the resulting β^{vt} is base-country independent.

Data

We base our analysis primarily on the Penn World Table (PWT), version 10.0, see Feenstra et al. (2015) for a general discussion of this dataset and www.ggdc.net/pwt for information on this most recent release. Most importantly, PWT incorporates all global ICP PPP comparisons since the first one for 1970 and up to the latest version for 2017. Country coverage has increased substantially over this period, from 16 in 1970 to 175 in 2017, see Table 13.1.[7] The statistical project has also become a much broader exercise, building on a growing body of knowledge regarding both conceptual and practical concerns when comparing prices across countries (World Bank, 2013). In ICP, GDP is built up from the expenditure side of the National Accounts, which means prices are collected for products used for household consumption, for government consumption and for investment. In PWT, estimates for prices of exported and imported products are added to get a complete accounting of GDP.[8] This means that, in addition to measuring the inconsistency between GDP PPPs and changes in the GDP deflator, we can also measure inconsistency at the level of the major expenditure categories.

Table 13.1 lists all global benchmarks included in PWT, the number of countries participating in each comparison and the number of countries that can be compared across different benchmarks. So, the 'vs. t-1' column shows the number of countries that were in both the current and the previous benchmark (so 16 in both ICP 1970 and ICP 1975), column

[7] 176 countries participated, but this includes Bonaire for which complete GDP-level data is not available, see World Bank (2020).

[8] In ICP the exchange rate is used to convert exports and imports to a common currency. PWT relies on the estimates by Feenstra and Romalis (2014) for estimates of quality-adjusted export and import prices.

Table 13.1 The number of participating countries in each ICP benchmark and the number of countries when comparing to previous benchmarks

ICP Benchmark	*Participating countries*	*vs. t-1*	*vs. t-2*	*vs. t-3*	*vs. t-4*	*vs. t-5*	*vs. t-6*	*vs. t-7*
1970	16							
1975	33	16						
1980	60	27	14					
1985	63	41	26	14				
1996	115	59	51	30	13			
2005	145	99	55	53	32	16		
2011	177	142	110	63	59	32	16	
2017	175	173	140	109	63	58	32	16

Notes Column 'vs. t-1' lists the number of countries that participated in both that comparison and the previous one, so 16 countries participated in ICP 1970 and those same 16 also participated in ICP 1975. Column 'vs. t-2' compares to two comparisons earlier, so only 14 of the countries that were in ICP 1970 were also in ICP 1980

'vs. t-2' shows the number that were in the benchmark and the benchmark before that (so 14 in both ICP 1970 and ICP 1980) and so on. It is good to note here that 1996 was not an official global ICP benchmark, but rather a synthetic one constructed for PWT versions 6.x based a PPP benchmark for a set of regions in 1993, linked to the OECD/Eurostat benchmark for 1996 (see Heston et al., 2002). And while ICP 2005 is a regular global benchmark, PWT corrects for methodological differences with the subsequent benchmarks and the bias in the linking of the regional comparisons (Deaton & Aten, 2017; Inklaar & Rao, 2017).

Table 13.1 illustrates that simply comparing the maximum set of countries across benchmarks leads to very unbalanced samples, with more countries covered in more recent years. Especially when trying to establish whether inconsistency has decreased over time, this sample variation can be problematic. But a balanced panel that covers all 7 ICP benchmarks would include no more than 13 countries. To strike a middle ground, we define a balanced sample using the 52 countries that participated in every ICP comparison since 1985.

Table 13.2 shows summary statistics for both samples. As the table illustrates, the balanced sample becomes less representative over time, as benchmarks after 1985 participation grew in particular among lower-income countries. Of further interest is that there is no clear trend in the average price level (column 2–3 and 6–7) and an increasing trend

Table 13.2 Summary statistics for every benchmark year for the full sample of countries and a balanced sample

Benchmark	*Full sample*				*Balanced sample*			
	(1) N	*(2)* p_Y	*(3)* p_C	*(4)* Y	*(5)* N	*(6)* p_Y	*(7)* p_C	*(8)* Y
1970	16	0.65	0.70	10,404				
1975	33	0.79	0.79	10,468				
1980	60	0.95	0.98	10,065				
1985	63	0.64	0.67	10,953	52	0.63	0.68	11,555
1996	115	0.73	0.76	13,162	52	0.80	0.83	16,546
2005	145	0.56	0.60	17,182	52	0.71	0.74	22,544
2011	177	0.67	0.68	21,470	52	0.78	0.80	24,836
2017	175	0.59	0.60	21,129	52	0.63	0.65	27,205

Notes This table shows descriptive statistics for the full sample and the balanced sample, which includes only countries that participated in every ICP benchmark since 1985. Shown are the number of countries N, the average price level for GDP (the PPP divided by the exchange rate) p_Y, the average price level of consumption p_C and the average GDP per capita level. The price levels are equal to 1 for the United States in every year. The GDP per capita level is in 2017 US dollars ($CGDP_o$/POP from PWT 10.0)

in the average income level. The main outlier is the ICP 1980 benchmark, which shows higher relative prices and lower-income levels than the 1975 or 1985 benchmark. This prefaces one of our findings, namely that the 1980 benchmark was the only one to substantially distort the international income distribution.

Most of the questions we ask of the data can be answered at this high level of aggregation. But to establish whether inconsistency is larger for harder-to-measure product categories, we also use more detailed information. Part of the release of the ICP 2017 results (World Bank, 2020) were PPPs for 2011 and 2017 at the so-called basic-heading level. Within household consumption, we can distinguish spending on food and non-alcoholic beverages (COICOP 01). Going one step more detailed is spending on food (011), on bread and cereals (0111) and, finally, on the basic-heading rice (01111).[9] Matched to this categorisation is information on inflation. Nearly all countries publish consumer price index (CPI)

[9] The aim of this statistical definition is to arrive at a fairly homogenous grouping of products. A practical consideration is that it is the lowest level of detail for which information about expenditure can still be compiled.

data at the two-digit COICOP level (e.g. food and non-alcoholic beverages) but some even at more detailed levels. In constructing a time-series of PPP for the period 2011 to 2017 (see Inklaar & Rao, 2020; World Bank, 2020) the most-detailed inflation series is allocated to each basic heading. That allows for the analysis of inconsistency at the basic-heading level between 2011 and 2017.

Results

We now turn to answering the questions we set out in the introduction, to assess patterns in the inconsistency data. As a starting point, it is helpful to gauge the overall size and scope of inconsistencies. Across all ICP benchmarks (comparing each benchmark to the next one), we find that the average inconsistency of GDP PPPs, computed as in Eq. (13.7), is −0.043. The variation in inconsistency over all countries and years is large with a range between the 25th and 75th percentile of [−0.148,0.071] and a range from the 5th to the 95th percentile of [−0.436,0.320]. This leads to a large average RMSI, based on Eq. (13.8), of 0.22. Even without our more detailed analysis that is to come, these descriptive statistics mean that caution is in order, especially when comparing countries that are close in income level as the size of inconsistencies make reversals of income rankings from one ICP benchmark to the next quite possible.

Are More Recent Global Benchmarks More Consistent?

Our first question is whether methodological improvements and more extensive resource allocation to statistical programmes have decreased inconsistency between more recent ICP benchmarks compared to earlier benchmarks. An alternative possibility is that the rising number and greater diversity of countries covered have made relative price estimations more difficult. Figure 13.2 shows RMSI estimates for consecutive benchmarks for GDP PPPs, distinguishing RMSI for the full sample and the balanced sample of countries. For the full sample, inconsistencies have increased and then decreased. The increase in inconsistency after the early benchmarks (1970–1985) is remarkable but we should emphasise that country coverage expanded substantially after these early benchmarks. Trying to compare prices across a more disparate group of countries is more difficult, as we discuss below. The decrease in inconsistency since 1985 is similar for the full sample and the balanced sample,

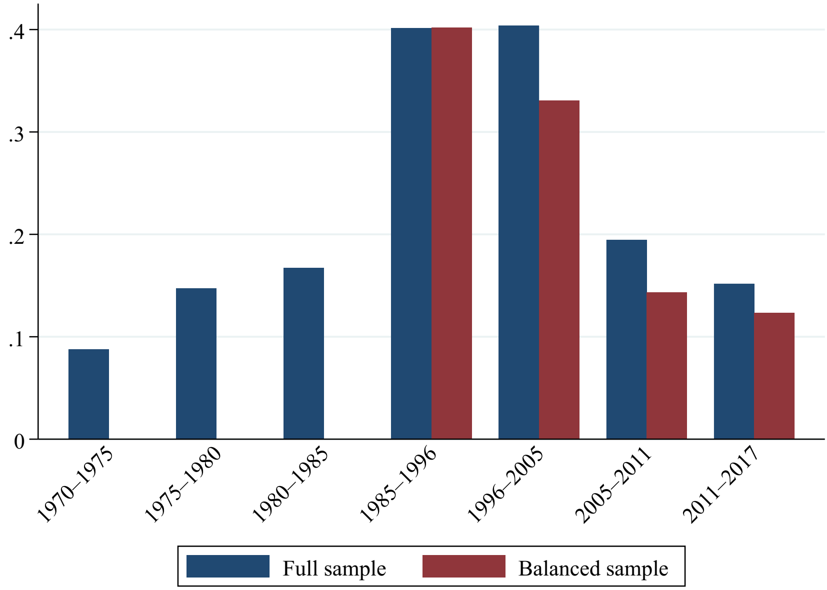

Fig. 13.2 RMSI between consecutive ICP benchmarks for GDP PPPs (*Notes* Each bar shows the root mean squared inconsistency (RMSI), calculated as in Eq. [13.8]. The inconsistency is computed between two consecutive PPP benchmarks, so the bar labelled '1970–1975' computes the inconsistency [using Eq. (13.7)] between the PPPs from ICP 1975 and those from ICP 1970. The full sample includes all countries that participated in the two consecutive PPP benchmarks [see Table 13.1 for the sample sizes]; the balanced sample includes the 52 countries that have participated in every ICP benchmark comparison since 1985.)

which provides support for the hypothesis that statistical improvements have decreased inconsistency.[10]

[10] The largest RMSIs involve the 1996 PPP benchmark (1985–1996 and 1996–2005), which is the only one that is not a proper global price comparison, see Heston et al. (2002). The RMSI for 1985–2005 is 0.29 for the full sample (see Table 13.4), which is notably lower than the ~0.4 in Fig. 13.2 for 1985–1996 and 1996–2005, but still higher than subsequent comparisons.

In Fig. 13.3, the degree of inconsistency over time is shown for each of the five major expenditure categories: household consumption expenditure, gross capital formation, government consumption expenditure, exports and imports. We use data for the full sample of countries and zoom in on the period since ICP 1985 as country coverage becomes broad enough for a meaningful comparison. The PPPs for the domestic expenditure categories (household consumption, investment and government consumption) are based on ICP benchmarks, the PPPs for exports and imports are introduced separately in PWT 10.0 and are estimating using trade unit value data for merchandise trade and following the methodology introduced by Feenstra and Romalis (2014).[11]

The figure shows a very comparable trend for the expenditure categories based on ICP benchmarks, mirroring the GDP-level trend from Fig. 13.2. By the final comparison, between ICP 2011 and ICP 2017, the inconsistency for household consumption is, at 0.13, substantially lower than for investment (0.33) or government consumption (0.28) and even lower than for exports (0.27) or imports (0.17). The decline in inconsistency is more marked for investment and government consumption, though. The decline for the ICP-based categories is monotonic, suggesting a continuous improvement in statistical practice, with particularly strong improvements since ICP 2005. That round marked the start of greater investment of resources by statistical agencies and it had, for the first time, the World Bank in its role as the host organisation for coordinating these efforts.

Comparing the downward trend in the RMSI for ICP-based categories with the much less pronounced trend for export and import PPPs is also informative. The data and methods for estimating export and import PPPs have been constant across these years, while ICP methods and data collection efforts have increased substantially. This is a further indication that it is improvements in PPP measurement that led to smaller inconsistency across recent rounds. The higher degree of inconsistency for investment and government consumption compared to household consumption may well be due to the greater prevalence of comparison-resistant expenditure categories, such as construction and collective consumption, a topic we return to in more depth, below.

[11] The Feenstra-Romalis type export and import PPPs are only available starting in 1984.

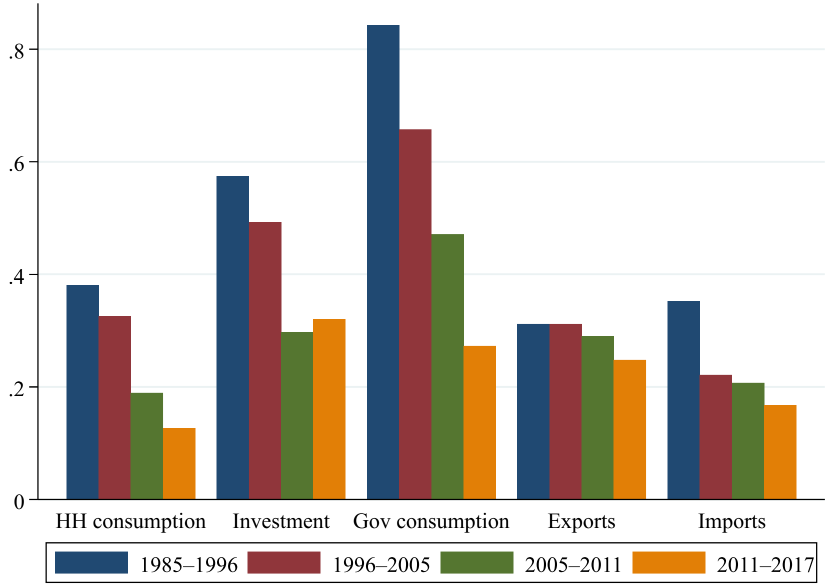

Fig. 13.3 RMSI between consecutive ICP benchmarks for PPPs by expenditure category since ICP 1985 (*Notes* Each bar shows the root mean squared inconsistency [RMSI], see also the notes to Fig. 13.2. The full sample of countries for each comparison is used. 'HH consumption' refers to the National Accounts expenditure category household consumption expenditure, 'Investment' to gross capital formation and 'Gov consumption' to government consumption expenditure. PPPs for these three expenditure categories are based on ICP benchmark data. PPPs for export and imports are from PWT 10.0 based on the method introduced by Feenstra and Romalis [2014] using data for merchandise trade, so excluding trade in services.)

Is There More Consistency Between More Similar Countries?

The second question we ask of the data is for which sets of countries inconsistency is a more prominent feature. The goal of the ICP is to make a global comparison of price and income levels, between the poorest and richest countries in the world. But, as stressed by Deaton and Heston (2010), comparing disparate countries involves the largest measurement challenges. This is for two reasons. First, when comparing two countries, index-number theory argues for the use of expenditure shares of both countries (see Eq. 13.5). Yet those expenditure shares are the outcome of consumer decision-making in each country and relative prices will shape spending patterns. Relying on only one country's expenditure shares for estimating relative prices will then impart a substitution bias. Using a Törnqvist index (as in Eq. 13.5) or a Fisher index[12] avoids substitution bias by using expenditure share information for both countries, but a consequence is that the comparison is made 'in the middle', i.e. reflecting neither country's spending pattern exactly. However, when expenditure shares are far apart, the 'comparison in the middle' might be a less accurate approximation. This could mean that country pairs with more dissimilar expenditure shares will have greater inconsistency.

A second reason for greater inconsistency for more disparate countries is that the set of products on which the PPP comparisons are based will be more dissimilar from the products included in country consumer price indexes (CPI). Consider two countries where in country 1 fish is the main source of (animal-based) protein and where meat is the more important source in country 2. In the extreme case, where country 1 consumes no meat and country 2 consumes no fish, it is not possible to even make a price comparison[13] but for country 1 the CPI will be based on fish prices and for country 2 it will be based on meat prices. In a more realistic case, where both countries consume both sources of protein, some variants of fish may be common in country 1 but not available (or only at relatively high prices) in country 2 and vice versa for some variants of meat. Within a product category, quality differences may lead to similar problems. Housing is a prime example, since a typical house or apartment in a high-income country may be very uncommon in a low-income country and vice versa.

To assess the importance of disparity between countries, we consider two indicators. The first is the (squared) difference between the expenditure shares of countries and we expect that a country pair with a larger difference between expenditure shares, $\delta s_{jk} = \sum_i (s_{ij} - s_{ik})^2$, will

[12] The Fisher index is the geometric mean of the Laspeyres price index, which compares prices using reference-country expenditure shares, and the Paasche price index, which uses comparison-country expenditure shares.

[13] This extreme outcome is also due to the two-country setup. If country 3 consumes meat and fish, an indirect comparison can be made between country 1 and 2 via country 3.

exhibit higher inconsistency, d_{jk}. That most closely tests the first reason, where inconsistency arises from index-number challenges. The second is to compute the (squared) difference in income levels and there we expect that a country pair with a larger difference, $\delta y_{jk} = (y_j - y_k)^2$, shows higher inconsistency. While imperfect, this proxy may capture aspects of both reasons.

Table 13.3 shows the results of these analyses for GDP PPPs and PPPs for household consumption. We compute the correlation by pair of ICP benchmarks, so, for example, the first row is based on comparing the PPPs from ICP 1970 and ICP 1975. The table shows that inconsistency is systematically higher for country pairs where expenditure patterns differ more, which is in line with prior expectations. Country pairs that differ more in income level also typically show higher inconsistency, but the evidence is less consistent. The correlations are also not large, less than 0.20 for more recent benchmarks. The general pattern of inconsistency thus supports the concerns by Deaton and Heston (2010) that comparisons of more disparate countries are more difficult. Yet the low correlations, especially for more recent benchmarks, indicate that our measures for approximating this disparity are imperfect and/or other (possibly random) factors contribute substantially to inconsistency as well.

Figure 13.4 provides another perspective on the question whether inconsistency is greater when comparing more disparate countries. For each pair of benchmarks, the countries that participated in both are ranked by average income level over the two years. The RMSI is then computed over each quartile of the sample. The figure shows that the RMSI was substantially greater for lower-income countries for the earlier ICP benchmarks, but this pattern is much more muted when comparing ICP 2005 and ICP 2011 and has even disappeared when comparing ICP 2011 and ICP 2017. So, while Table 13.3 illustrates that inconsistency is a larger concern when income differences are large and/or when expenditure patterns are very different, improvements in price measurement seem to have helped reduce inconsistencies, primarily for lower-income countries. From this analysis, we cannot conclude whether that is due to improved price sampling and measurement for PPPs, for CPIs or a closer alignment of PPP and CPI price samples.

Table 13.3 Correlation between inconsistency and expenditure share differences and income level differences for GDP and household consumption

	GDP		*Household consumption*	
	Expenditure share	*Income level*	*Expenditure share*	*Income level*
1970–1975	0.45*	−0.04	0.44*	−0.06
1975–1980	0.52*	0.21*	0.41*	0.16*
1980–1985	0.27*	0.47*	0.11*	−0.03
1985–1996	0.22*	0.10*	0.20*	0.06*
1996–2005	0.22*	0.04*	0.07*	0.03*
2005–2011	0.21*	0.08*	0.23*	0.03*
2011–2017	0.20*	0.00	0.15*	0.10*

Notes For each pair of subsequent ICP comparisons (see notes to Fig. 13.2), the inconsistency d_{jk} is computed for all country pairs for GDP PPPs and household consumption PPPs. That inconsistency is correlated with the squared difference between expenditure shares, $\delta s_{jk} = \sum_i (s_{ij} - s_{ik})^2$, and with the squared difference in income level $\delta y_{jk} = (y_j - y_k)^2$. The income level is the (PPP-converted) GDP per capita level in 2017 US dollars ($CGDP_o$/POP from PWT 10.0), averaged over the two comparisons. Since there are 16 countries that participated in both ICP 1970 and ICP 1975 (see Table 13.1), the correlations in the first row are based on $\frac{1}{2} \times 16 \times (16 - 1) = 120$ observations. * denotes a correlation coefficient significantly different from zero at the 5-per cent level

Is There More Consistency for Products that Are Easier to Price and Compare Across Countries?

As already seen in Fig. 13.3, different expenditure categories show larger inconsistencies than others. Zooming in on more detailed expenditure categories, we know that some are considered 'comparison-resistant' such as construction and collective services. In addition, PPPs for some detailed expenditure categories do not rely on direct price observations. Instead, in ICP these are based on PPPs from other categories or higher aggregates; these are referred to as reference PPPs. For example, rather than being directly observed, the PPP for narcotics is based on the PPPs for pharmaceutical products and for tobacco.

To assess whether there is more consistency for expenditure categories with items that are easier to price and compare across countries, we compute the RMSI for the 12 expenditure categories within household consumption expenditure between ICP 2011 and ICP 2017. These are the two rounds of ICP for which we have both PPPs for detailed expenditure categories (basic headings) and inflation series at a more specific level of detail than an overall CPI. Limiting this comparison to household

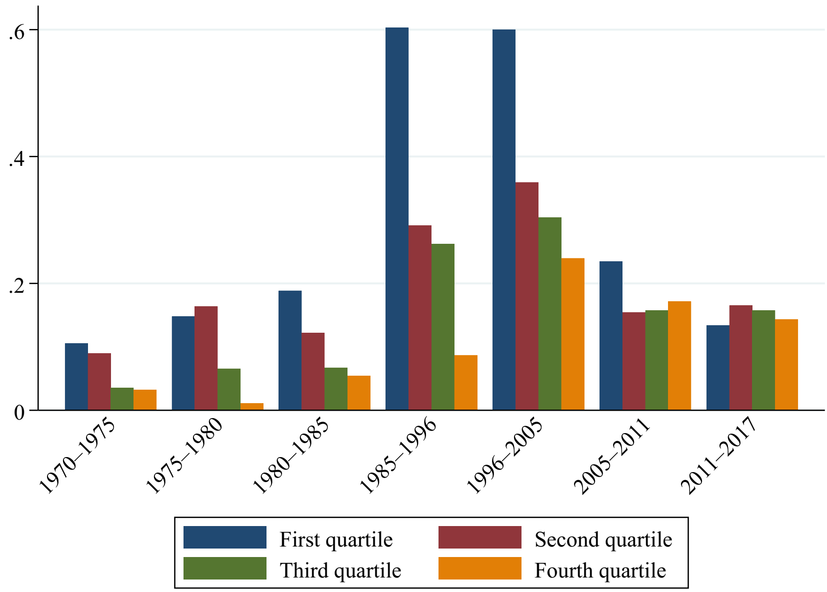

Fig. 13.4 RMSI for GDP PPPs by income quartile (*Notes* Figure shows the root mean squared inconsistency [RMSI] across consecutive ICP benchmarks, see also notes to Fig. 13.2. For each comparison, the countries are split into quartiles by income level. The income level is computed in each year as GDP per capita [$CGDP_o$/POP from PWT 10.0] relative to the United States and that relative position is averaged over the two years of the PPP benchmarks. These calculations are based on the unbalanced sample of countries)

consumption expenditure categories is because these are the categories for which direct PPP measures are most commonly available and because CPI data is typically available at this level. We could also make this comparison for other expenditure categories, but the interpretation is more difficult because there is a larger difference between the approach taken for measuring PPPs and the approach for measuring inflation. For example, ICP relies on the exchange rate for the balance of exports and imports while national statistical agencies would construct export and import price indexes.

In Fig. 13.5, we show the unweighted average of RMSI estimates within each expenditure category.[14] We make a somewhat ad-hoc distinction between categories where price measurement is relatively hard (the red bars) and where it is easier (blue bars). This distinction is based on methodological considerations, as in World Bank (2013), where challenges for measuring the relative price of housing or education are discussed at length. The 'hard-to-measure' categories also more frequently rely on reference PPPs rather than direct price measurement.

The figure suggests some relationship between the RMSI and whether a category is harder or easier-to-measure. An easier-to-measure category such as food has a low RMSI, while education, housing and health show a higher RMSI. At the same time, the communication category has the highest RMSI, and the furnishing and household equipment category has a higher RMSI than health. A binary easy/hard distinction is not ideal as there are gradations in measurement challenges, even within these broad categories.[15] Yet the current distinction does suggest that inconsistency is a more substantial problem where measurement problems are thornier.[16] We also compared the average RMSI for basic-heading categories based on reference PPPs and those based on direct price measurement, which is another perspective on easy versus hard-to-measure. The average RMSI for both groups is very similar with the direct price measurement categories even showing a slightly higher RMSI (0.35) than those based on reference prices (0.32). So, while Fig. 13.5 indicates that more serious PPP measurement problems can lead to more substantial inconsistency (and thus that improving measurement in those areas may help reduce inconsistency), the pattern of inconsistency is more varied than a simple easy-vs-hard to measure distinction.

[14] The figure is very similar when weighting by expenditure shares within each category or when computing RMSI at the level of aggregate PPPs and corresponding CPIs. For both these alternative approaches, the hard-to-measure categories are at or near the top of the ranking as in Fig. 13.5.

[15] For example, PPPs for (imputed) housing rentals are among the more challenging to measure, but PPPs for electricity are conceptually much easier.

[16] When considering all basic-heading categories within these broader categories as either 'easy' or 'hard' to measure, the RMSI for the 'hard' categories is 0.44 versus 0.33 for the 'easy' categories, a statistically significant difference.

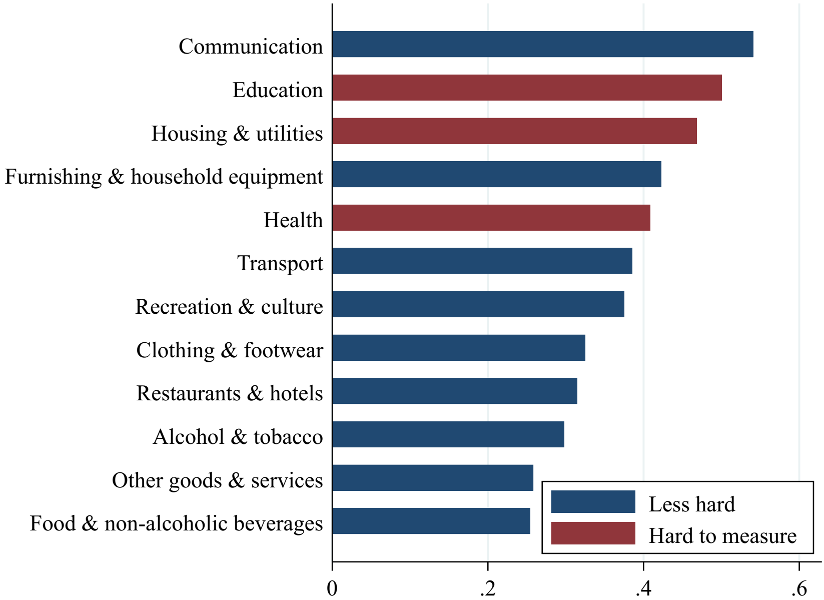

Fig. 13.5 RMSI by household consumption expenditure category between ICP 2011 and ICP 2017 (*Notes* Root mean squared inconsistency [RMSI] is computed at the basic-heading level using data from ICP 2011 and ICP 2017. An unweighted average of these basic-heading level RMSIs within each household consumption expenditure category is computed and shown here. Blue bars ('less hard') are those categories for which price measurement challenges are modest; red bars ['hard-to-measure'] cover categories that are sometimes described as 'comparison-resistant'. There, products and price-defining characteristics are hard to define precisely or where no direct pricing is feasible)

Would More Frequent Benchmark Comparisons Lead to More Consistency?

A typical response by statisticians to inconsistencies between PPP benchmarks is to increase the frequency of cross-country price comparisons. Indeed, Eurostat provides annual estimates of PPPs based on a rolling-survey approach where every expenditure category is covered once every two or three years. This ensures that PPPs need not be estimated solely based on inflation information. Increasingly, regional agencies, who coordinate efforts of statistical agencies in different regions of the world, are moving to more frequent cycles. For example, Western Asia's ESCWA has been publishing annual PPPs since the results for 2011. The ICP is also accelerating its benchmark cycle from 6 years (since 2005) to every 3 years—though due to the COVID-19 the planned ICP round for 2020 was shifted to 2021.

There are broadly two possibilities for what the effect of such more frequent benchmarks could be. First, if there are systematic reasons for inconsistency, such as the term based on differing expenditure shares in Eq. (13.6), it could be that inconsistencies accumulate over time and that shortening the time period between benchmarks would lead to a smaller RMSI. Alternatively, if the reasons for inconsistency are largely random and caused by variation in product samples and in inflation rates between products, then it could be that inconsistency remains stable. Such a result would not imply that more frequent benchmark comparisons are useless. More frequent comparisons might be useful to maintain expertise in statistical agencies, be relatively cheaper because surveys need not be setup anew and help to track changing product availability more closely.

This question is also relevant for historical price and income comparisons. Maddison (2001, 2007) relied on a single modern benchmark comparison of income levels for the year 1990 and then used data on GDP per capita growth to extend these figures back in time to the nineteenth century and earlier.[17] This approach was certainly defensible when Maddison first developed his data, but since then there have been increasing efforts to develop contemporaneous income comparison. A good example is the recent work by Ward and Devereux (2021), providing PPP estimates for a set of economies for 1872 and 1910, but see also Bolt et al. (2018) for an overview of such studies. The greater availability of historical income comparisons raises the question how to assess the inconsistency between those historical figures and the original Maddison figures (or other projections from modern price comparisons). If inconsistency tends to accumulate over time, for instance, because consumption has dramatically shifted from food to services, the projections of GDP per capita data by Maddison (2001, 2007) over very long time periods are much harder to defend. If random variation dominates, inconsistency could still be substantial, but it would be harder to discount

[17] This 1990 benchmark does not correspond to a single ICP PPP benchmark, but is instead based on a variety of sources, including ICP 1985, ICP 1980, Eurostat and OECD comparisons and (via PWT) price comparisons based on expat cost-of-living indexes.

estimates based on modern price comparisons, especially when these are based on higher-quality data than can be used in historical comparisons.

Though it is not possible to assess how moving from a 6-year to a 3-year benchmark cycle would affect inconsistency moving forward, we can look back and assess how a longer time period between benchmarks would have affected inconsistency. This also brings us closer to the time frames for historical income comparisons. Table 13.4 shows what happens to the RMSI when not comparing consecutive benchmarks. The first row shows the RMSI for the balanced panel of 52 countries since ICP 1985 based on consecutive ICP benchmarks, so our approach so far. The top-left figure of 0.40 is the RMSI when comparing ICP 1985 to ICP 1996 and is the same as shown in Fig. 13.2. The second row skips one benchmark, so the first figures in that row (0.29) is the RMSI when comparing ICP 1985 to ICP 2005, the second figure (0.37) when comparing ICP 1996 to ICP 2011, and so on.

Reading this table down the diagonal, so with the same initial ICP benchmark but skipping more benchmarks, does not show a clear trend in the RMSI. For ICP 1985 as a starting point, going to 1996 leads to a larger RMSI (0.40) than skipping to 2005 (0.29), 2011 (0.32) or 2017 (0.33). Starting from ICP 1996 shows a small increase in the RMSI, from 0.33 via 0.37 to 0.35. Another perspective is going by row and then the average RMSI for '1 benchmark apart' is lower than for 2, 3 or 4 benchmarks apart. However, there is no clear difference between 3 or 4 benchmarks apart, so even if inconsistency were to increase with longer

Table 13.4 RMSI across multiple benchmarks

	Final benchmark			
	1996	*2005*	*2011*	*2017*
1 Benchmark apart (baseline)	0.40	0.33	0.14	0.12
2 Benchmarks apart		0.29	0.37	0.17
3 Benchmarks apart			0.32	0.35
4 Benchmarks apart				0.33

Notes The table shows the root mean squared inconsistency (RMSI) for GDP PPPs using the balanced dataset for 52 countries. Each row shows the interval between benchmarks, with the first row showing the baseline figures with inconsistency computed based on subsequent benchmarks, so the top-left figure (0.40) is the RMSI when comparing ICP 1985 to ICP 1996. The second row is based on two benchmarks apart, so the first figure (0.29) is based on comparing ICP 1985 to ICP 2005, the second (0.37) based on comparing ICP 1996 to ICP 2011, and so on time periods between benchmarks, it is not clear whether that trend would continue. So, conversely, whether inconsistency would decrease if ICP benchmarks become more frequent is uncertain. With the relatively small number of ICP benchmarks, though, caution is in order in drawing conclusions. Caution is warranted even more because of the variable number of years between ICP benchmarks, with an 11-year gap between 1985 and 1996 and a 9-year gap between 1996 and 2005 but 6-year gaps since then.

Drawing conclusions relevant for historical income comparisons is also hazardous since the time frames are even longer. This will lead to larger differences in economic structure and spending patterns. Disruptions such as the World Wars may hamper reliability of statistics over time even more than in current times, but the lack of high-quality statistics going back further in time may also raise doubts about the quality of historical income comparisons.

Do Inconsistencies Distort the International Income Distribution?

Inconsistency between PPP benchmarks and relative inflation is problematic when trying to assess the relative income level of individual countries, but it is even more worrisome when the entire income distribution changes as a result. This was a main concern when ICP 2011 was released and international income differences were notably smaller than had been expected based on ICP 2005 PPPs that were extrapolated using relative inflation rates (Deaton & Aten, 2017; Inklaar & Rao, 2017), i.e. income levels of low-income countries were closer to those of high-income countries than had been expected. In the context of Eq. (13.9), this meant that low-income countries had predominantly negative inconsistency estimates since a negative inconsistency on PPPs implies a positive inconsistency on real income levels. High-income countries were on average closer to zero.

Table 13.5 shows the regression coefficients on income levels based on inconsistency in subsequent ICP benchmark years. Recall that for ICP 2005, we use the PPPs that are part of PWT 10.0 based on the adjustments proposed by Inklaar and Rao (2017) to address the distortions that the original regional linking procedure had imparted. As the table shows, the coefficient for 2005–2011 is not significantly different from zero, which corresponds to the result of Inklaar and Rao (2017). Indeed, the only coefficients that are significantly different from zero are the two involving ICP 1980. The inconsistency between ICP 1975 and

Table 13.5 Regression coefficients of GDP PPP inconsistency on income levels

Benchmarks	*Coefficient*	*s.e*	*# of countries*
1970–1975	0.039	(0.019)	16
1975–1980	−0.092*	(0.021)	27
1980–1985	0.100*	(0.020)	41
1985–1996	−0.095	(0.054)	59
1996–2005	−0.033	(0.040)	99
2005–2011	−0.025	(0.014)	142
2011–2017	0.014	(0.010)	173

Notes The table shows coefficient estimates of β from Eq. (13.9), so the extent to which countries with higher income levels show greater inconsistency for GDP PPPs. The income level is the level of GDP per capita ($CGDP_o$/POP from PWT 10.0), averaged between the two benchmarks, relative to the United States. Robust standard errors are in parentheses in the column 's.e'. * denotes a coefficient significantly different from zero at the 5-per cent level

ICP 1980 is negatively related with income level, which implies that international income differences were unexpectedly larger based on the 1980 PPP estimates than based on the 1975 PPP estimates. Going from 1980 to 1985, this pattern was reversed with a positive and significant coefficient of similar size as before. The lack of systematic inconsistencies is comforting and implies that, despite the large inconsistencies for individual countries, the broad cross-country pattern of income differences is typically not distorted.

Conclusions

The topic of this paper is the quality of cross-country price and income comparison benchmarks. In our perspective on this topic, we are close to Rao et al. (2010), who build a statistical model to reconcile inconsistencies between different benchmarks and information about inflation and estimate a 'consistentised' real income series. But rather than reconciliation, our aim is to document patterns in inconsistency: has it increased or decreased over time? Is it larger for some comparisons and products than for others? With this exploration, we aim to provide more context to interpreting relative income estimates since the 1970s, point the way to where future measurement efforts could be most fruitfully applied, and provide a better understanding of patterns in the data to help underpin more statistical efforts.

One conclusion we draw is that it is likely that improved statistical methods for measuring PPPs have decreased inconsistency. Inconsistency based on ICP PPP data has decreased over time, most markedly since ICP 2005, which saw major investments of resources and methodological improvements. Inconsistency has not changed notably for export and import PPPs, which were estimated based on the same data and methods for the entire period, which is further support for our conclusion. Most of the measurement gains were made in comparing income levels of low-income countries, which has no doubt improved our ability to trace global income inequality and put a firmer basis under the World Bank's figures for absolute poverty. In a further reassuring result, we show that—with the exception of ICP 1980—inconsistency has not shifted the international income distribution. This means that low-income countries were as likely to see their income level relative to high-income countries improve as deteriorate.

Yet inconsistency remains substantial, with a root mean squared inconsistency of 0.1–0.2. This implies that an adjustment in income levels of 10–20% is not uncommon when new PPP data are published. Inconsistency is lower when comparing countries with similar expenditure patterns and at more similar income levels. We also find that, within household consumption, inconsistency is higher in expenditure categories where PPP measurement challenges tend to be more substantial. This suggests that improved measurement methods in those areas could help reduce inconsistency even further.

Finally, we find that increasing the period of time between PPP benchmarks does not lead to larger inconsistency, which points to random variation in product sampling and inflation as a primary factor, rather than a systematic accumulation of inconsistency. This could mean that shortening the period between benchmarks would not lead to lower inconsistency. But while reduced inconsistency may not be an automatic outcome of more frequent international price comparisons, there are still good reasons to support these from a broader price measurement perspective. An important institutional argument is that maintaining the expertise that has been developed over the past 15–20 years in measuring PPPs is easier to maintain and extend when that expertise is more frequently called upon, since procedures remain operational and there will be more overlap between staff trained in these procedures. The PPP programmes run by Eurostat and OECD, that published PPPs at more frequent rates than the ICP, serve as key examples of such sustained expertise.

That institutional perspective is also helpful because there may be domestic spillovers from more frequent international price comparisons. Especially in countries with limited resources, the support from the ICP can help maintain and extend expertise in price measurement, which can also be put to good use in constructing more reliable CPI and other domestic price indexes. As discussed earlier, inconsistency between the change in PPPs and relative inflation need not mean that the PPPs are measured in error, it could be due to domestic inflation measurement problems or deficiencies as well.

A broader conclusion we draw from the analysis in this paper is how hard it still is to make international income comparisons, a conclusion shared with Deaton and Heston (2010). A point in case is the question whether China or the United States has the larger economy. In ICP 2017, the World Bank (2020) data show that the two countries were of approximately the same size, with a difference in GDP level of 0.5%. In the same data for 2011, the US economy was 10.7% larger than the Chinese economy. Yet given the size of the inconsistencies we discussed in this paper, a cautious person would have said the two economies were approximately the same size in that year as well.

References

Bolt, J., Inklaar, R., De Jong, H., Van Zanden, J. L. (2018). *Rebasing 'Maddison': New income comparisons and the shape of long-run economic development* (Maddison Project Working Paper 10).

Ciccone, A., & Jarociński, M. (2010). Determinants of economic growth: Will data tell? *American Economic Journal: Macroeconomics, 2*(4), 222–246. https://doi.org/10.1257/mac.2.4.222

Deaton, A. (2010). Price indexes, inequality, and the measurement of world poverty. *American Economic Review, 100*(1), 5–34. https://doi.org/10.1257/aer.100.1.5

Deaton, A., & Aten, B. (2017). Trying to understand the PPPs in ICP 2011: Why are the results so different? *American Economic Journal: Macroeconomics, 9*(1), 243–264. https://doi.org/10.1257/mac.20150153

Deaton, A., & Heston, A. (2010). Understanding PPPs and PPP-based national accounts. *American Economic Journal: Macroeconomics, 2*(4), 1–35. https://doi.org/10.1257/mac.2.4.1

Deaton, A., & Schreyer, P. (2021). GDP, wellbeing, and health: Thoughts on the 2017 round of the international comparison program. *Review of Income and Wealth*. https://doi.org/10.1111/roiw.12520

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review, 105*(10), 3150–3182. https://doi.org/10.1257/aer.20130954

Feenstra, R. C., & Romalis, J. (2014). International prices and endogenous quality. *The Quarterly Journal of Economics, 129*(2), 477–527. https://doi.org/10.1093/qje/qju001

Hajargasht, G., Hill, R. J., Rao, D. S. P., & Shankar, S. (2018). *Spatial chaining in international comparisons of prices and real incomes* (Graz Economics Papers 2018-03). University of Graz.

Heston, A., Summers, R., & Aten, B. (2002, October). *Penn World Table Version 6.1.* Center for International Comparisons at the University of Pennsylvania (CICUP).

Hill, R. J. (2004). Constructing price indexes across space and time: The case of the European Union. *American Economic Review, 94*(5), 1379–1410. https://doi.org/10.1257/0002828043052178

Hill, R. J., & Melser, D. (2015). Benchmark averaging and the measurement of changes in international income inequality. *Review of World Economics/Weltwirtschaftliches Archiv, 151*(4), 767–801. https://doi.org/10.1007/S10290-015-0229-6

Inklaar, R., & Rao, D. S. P. (2017). Cross-country income levels over time: Did the developing world suddenly become much richer? *American Economic Journal: Macroeconomics, 9*(1), 265–290. https://doi.org/10.1257/mac.20150155

Inklaar, R., & Rao, D. S. P. (2020, February 20–21). *ICP PPP time series implementation*. 5th Meeting of the International Comparison Program (ICP) Technical Advisory Group (TAG), World Bank.

Inklaar, R., & Timmer, M. P. (2013). A note on extrapolating PPPs. In International Comparison Program (Ed.), *Measuring the real size of the world economy: The framework, methodology, and results of the International Comparison Program—ICP* (pp. 502–506). World Bank.

Johnson, S., Larson, W., Papageorgiou, C., & Subramanian, A. (2013). Is newer better? Penn World Table Revisions and their impact on growth estimates. *Journal of Monetary Economics, 60*(2), 255–274. https://doi.org/10.1016/j.jmoneco.2012.10.022

Krijnse Locker, H., & Faerber, H. D. (1984). Space and time comparisons of purchasing power parities and real values. *Review of Income and Wealth, 30*(1), 53–83. https://doi.org/10.1111/j.1475-4991.1984.tb00477.x

Maddison, A. (2001). *The world economy: A millennial perspective*. OECD.

Maddison, A. (2007). *Contours of the world economy 1–2030 AD: Essays in macroeconomic history*. Oxford University Press.

McCarthy, P. J. (2013). Extrapolating PPPs and comparing ICP benchmark results. In: International Comparison Program (Ed.), *Measuring the real size of*

the world economy: The framework, methodology, and results of the International Comparison Program—ICP (pp. 473–506). World Bank.
Prados de la Escosura, L. (2000). International comparisons of real product, 1820–1990. *Explorations in Economic History, 37*, 1–41.
Rao, D. S. P., & Hajargasht, G. (2016). Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP). *Journal of Econometrics, 191*(2), 414–425. https://doi.org/10.1016/j.jeconom.2015.12.012
Rao, D. S. P., Rambaldi, A. N., & Doran, H. (2010). Extrapolation of purchasing power parities using multiple benchmarks and auxiliary information: A new approach. *Review of Income and Wealth, 56*(s1), S59–S98. https://doi.org/10.1111/j.1475-4991.2010.00386.x
Summers, R., & Heston, A. (1991). The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988. *The Quarterly Journal of Economics, 106*(2), 327–368. https://doi.org/10.2307/2937941
United Nations Statistical Commission. (1999). *Evaluation of the International Comparison Programme*. https://pubdocs.worldbank.org/en/164821487203245266/UNSC-30-Session-ryten-report-EN-1999.pdf
Veenstra, J. (2015). Output growth in German manufacturing, 1907–1936: A reinterpretation of time-series evidence. *Explorations in Economic History, 57*, 38–49. https://doi.org/10.1016/j.eeh.2015.03.001
Ward, M., & Devereux, J. (2021). New income comparisons for the late nineteenth and early twentieth century. *Review of Income and Wealth, 67*(1), 222–247. https://doi.org/10.1111/roiw.12466
Woltjer, P. J. (2015). Taking over: A new appraisal of the Anglo-American productivity gap and the nature of American economic leadership ca. 1910. *Scandinavian Economic History Review, 63*(3), 1–22.
World Bank. (2013). *Measuring the real size of the world economy: The framework, methodology, and results of the International Comparison Program—ICP*. World Bank. https://openknowledge.worldbank.org/handle/10986/13329
World Bank. (2020). *Purchasing power parities and the size of world economies: Results from the 2017 International Comparison Program*. World Bank. https://openknowledge.worldbank.org/handle/10986/33623