ROUTLEDGE ADVANCES IN RESEARCH METHODS

Meta-Regression Analysis in Economics and Business

T.D. Stanley and Hristos Doucouliagos

ROUTLEDGE

Meta-Regression Analysis in Economics and Business

Meta-Regression Analysis in Economics and Business is the first guide through the rapidly expanding field of meta-analysis in economics and business. Have you ever wondered, for example, whether a raise in the minimum wage really lowers employment or if taxes will cause people to conserve water? Meta-analysis is the way that science takes stock of our vast research output. Meta-analysis is a statistical and systematic review of all relevant research. It produces the authoritative assessments required for evidence-based practice in medicine, social sciences, economics, and business.

The purpose of this book is to introduce novice researchers to the tools of meta-analysis and meta-regression analysis and to summarize the state of the art for existing practitioners. Meta-regression analysis addresses the rising "Tower of Babel" that current economics and business research has become. Meta-analysis is the statistical analysis of previously published, or reported, research findings on a given hypothesis, empirical effect, phenomenon, or policy intervention. It is a systematic review of all the relevant scientific knowledge on a specific subject and is an essential part of the evidence-based practice movement in medicine, education, and the social sciences. However, research in economics and business is often fundamentally different from what is found in the sciences and thereby requires different methods for its synthesis—meta-regression analysis. This book develops, summarizes, and applies these meta-analytic methods.

Meta-Regression Analysis in Economics and Business offers the first comprehensive guide to conducting and understanding the type of meta-analysis (meta-regression analysis) needed for econometric studies. Actual systematic reviews of research are used throughout the book to illustrate the use of these meta-analytic methods. Among other things, it contains the first theory of meta-regression analysis, novel methods for correcting publication bias, and a rigorous demonstration that study quality will not affect meta-regression analysis.

T.D. Stanley is Bill and Connie Bowen Odyssey Professor of Economics at Hendrix College, Conway, AR, USA.

Hristos Doucouliagos is Professor in the School of Accounting, Economics and Finance, Deakin University, Melbourne, Australia.

Routledge Advances in Research Methods

1 E-Research
Transformation in scholarly practice
Edited by Nicholas W. Jankowski

2 The Mutual Construction of Statistics and Society
Edited by Ann Rudinow Sætnan, Heidi Mork Lomell, and Svein Hammer

3 Multi-Sited Ethnography
Problems and possibilities in the translocation of research methods
Edited by Simon Coleman and Pauline von Hellermann

4 Research and Social Change
A relational constructionist approach
Sheila McNamee and Dian Marie Hosking

5 Meta-Regression Analysis in Economics and Business
T.D. Stanley and Hristos Doucouliagos

Meta-Regression Analysis in Economics and Business

T.D. Stanley and
Hristos Doucouliagos

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2012
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Stanley, T. D., 1950–
Meta-regression analysis in economics and business/T.D. Stanley and Hristos Doucouliagos.
p. cm.
1. Economics–Research–Methodology. 2. Economics–Research–Evaluation–Statistical methods. 3. Economics literature–Evaluation–Statistical methods. 4. Regression analysis. I. Doucouliagos, Hristos. II. Title.
HB131.S73 2012
330.01’519536–dc23
2011052596

ISBN: 978-0-415-67078-4 (hbk)
ISBN: 978-0-203-11171-0 (ebk)

Typeset in Times New Roman
by Sunrise Setting Ltd

Contents

Figures

Tables

Acknowledgments

We are especially grateful for comments and suggestions from Margaret Giles, Jost Heckemeyer, Julian Higgins, Stian Skår Ludvigsen, Debdullal Mallick, Jon Nelson, Geoff Pugh, Randy Rosenberger, and Hossam Zeitoun. Our research collaborators over the years have been instrumental in the development of our ideas: Janto Haman, Steve Jarrell, Patrice Laroche, Martin Paldam, Andrew Rose, Randy Rosenberger, and Mehmet Ulubasoglu. Furthermore, we need to acknowledge the ideas, feedback and support that we received from numerous scholars during various seminars and MAER Network colloquia. Needless to say, any errors or omissions are solely our responsibility.

1 Introduction

> This is but the start of their undertakings! There will be nothing too hard for them to do. Come, let us go down and confuse their language on the spot so that they can no longer understand one another.
>
> (*Genesis* 11: 6–7)

1.1 The Tower of Research

We live in a wondrous age. Information technology has given billions access to the world's accumulated scientific knowledge as well as this week's viral video of some kid dancing. Inexpensive hand-held devices bring us the contents of a hundred libraries in seconds and the processing power of the best computers from only a generation ago. But has society's knowledge become thousands of miles wide and mere nanometers deep? To some, these gigabytes, terabytes, and petabytes usher in a Renaissance of human knowledge and creativity. To others, like the Nobel econometrician, James Heckman, they represent a tsunami of noise and misinformation that threatens to drown out genuine scientific knowledge and informed policy action (Heckman, 2001). How will we be able to distinguish useful information from mere exaggeration, ideology and even lies?

Our extraordinary era has seen the rapid expansion of research publications, the meteoric rise in empirical economics and business research, and the proliferation of increasingly narrow areas of academic research. Is this not another "Tower of Babel," one where these terabytes ensure "that [we] can no longer understand one another"?

Worse than the sheer mass of information are the large differences in what researchers report about a given phenomenon, treatment or effect. In social science, economics, and business research, one always finds a large variation in the reported estimates of a given parameter. The rising pressure to publish, with its concomitant demand to uncover something novel, is sufficient to generate ample conflict among empirical findings. Because economics and business ultimately depend on human behavior, the empirical phenomena that we study will always contain a great deal of natural variation; that is, a genuine heterogeneity that depends on prevailing socio-political institutions and history. Often, it seems as if researchers speak different languages.

Incentives in the media, science, and the academy all seem to accentuate the dissidence of reported research. Should science become too clear and uncontroversial (e.g. the health effects of smoking, global warming or evolution), concerned groups will fund researchers and spokesmen to manufacture uncertainty and controversy. Yet, wide variation in research findings will likely occur without any outside intervention. Even the best scientific practice will produce very disparate research findings without resorting to anything ethically questionable. Science progresses through critical discourse and by challenging what is believed. When virtually all researchers agree about a given theory, empirical phenomenon, or policy effect, scientific progress is likely to stagnate, and, ironically, we find larger, more distorting biases in what researchers report (Doucouliagos and Stanley, 2012).[1]

Although we need not fear disparate scientific findings, practical policy demands clarity. Without some intelligent summary of business and economic research, understanding and informed policy actions are impossible. Yet, conventional narrative reviews are fatally flawed. Because there are no objective standards, conventional reviewers often dismiss studies or findings that do not fit into their preconceived notions or theories (Stanley, 2001). "Believing is seeing" (Demsetz, 1974: 164). Beliefs are often self-fulfilling. One can almost always find research papers or a literature review that interprets past research through the reader's own priors or ideological lens. Yet, without the reliable coherence that a good narrative review is meant to provide, conflicting research results overwhelm any clear understanding of economic phenomena. The only informed and correct conventional summary of the research record on nearly any important economic phenomenon or policy question is: "it depends."

What we need is some objective and critical methodology to integrate conflicting research findings and to reveal the nuggets of "truth" that have settled to the bottom. Meta-regression analysis (MRA), when replicable and conducted properly, offers such methodology. We believe that it is economics' best hope for genuine empirical progress.

Meta-analysis is the statistical analysis of previously published, or reported, research findings on a given hypothesis, empirical effect, phenomenon, or policy intervention. It is a systematic review of all the relevant scientific knowledge on a specific subject and is an essential part of the "evidence-based practice" movement in medicine, education and the social sciences.

Medical researchers have long embraced meta-analysis to provide an objective and comprehensive summary of the often conflicting results from randomized clinical trials (RCTs) of some drug or medical procedure. Evidence-based medical practice has changed how sick people are treated and saved 100,000 lives within the first 18 months of its adoption (Berwick *et al*., 2006; Ayers, 2007). Because RCTs tend to be very expensive and time-consuming, medical practice is often based on only a few trials, trials which often report conflicting success and risks. To economize on this limited and expensive scientific evidence, medical researchers have been employing meta-analysis for over 30 years (Chalmers *et al*., 1977). Often, when several RCTs are statistically combined, a clearer, more accurate picture of a given treatment's efficacy emerges.

Meta-analysis is the most objective and statistically rigorous approach to systematic reviews, which, in turn, provides the evidence for the evidence-based practice movement. A systematic review differs from more conventional narrative reviews by conducting exhaustive searches in a serious attempt to include all studies meeting explicitly stated criteria. When conducted properly, a systematic review is replicable by independent reviewers.

In economics, meta-analysis is almost entirely meta-regression analysis, and it has a somewhat different focus than how it is applied in other fields. MRA was initially proposed to correct known misspecification biases, endemic among econometrics estimates (Stanley and Jarrell, 1989). Meta-regression analysis is a multivariate empirical investigation, using multiple regression analysis, of what causes the large variation among reported regression estimates or transformations of regression estimates (e.g. elasticities, environmental values, or partial correlations). Because econometrics is typically observational (i.e. non-experimental), even the most rigorous econometric applications cannot eliminate all the potentially confounding influences.[2] By now, hundreds of MRAs have confirmed that such misspecification biases are routinely found in all areas of empirical economics research, and many of these are large enough to have a significant practical effect on how we view the phenomenon in question or on how a given policy intervention is evaluated.

Then there is the question of selection. Only a few of potentially millions of econometric models are reported – "I just ran two million regressions" (Sala-i-Martin, 1997).

> Empirical results reported in economics journals are selected from a large set of estimated models. Journals, through their editorial policies, engage in some selection, which in turn stimulates extensive model searching and prescreening by prospective authors. Since this process is well known to professional readers, the reported results are widely regarded to overstate the precision of the estimates, and probably to distort them as well. As a consequence, statistical analyses are either greatly discounted or completely ignored.
>
> (Leamer and Leonard, 1983: 306)

Each of these model specification choices affects the reported results, often by a lot, and there is no reliable way to know which model specification is correct.[3] Enter meta-regression analysis.

Meta-regression analysis can explicitly model the effects of observed model specification variation and thereby directly estimate the associated misspecification biases. Accommodating and correcting the biases associated with applied econometrics is the central objective of MRA. Meta-regression analysis is a systematic and comprehensive review of all existing, yet comparable, empirical evidence. It allows the systematic reviewer to model and estimate any explanatory or biasing factor for which information or a proxy is available and thereby filters out their influence on our scientific knowledge. This applies to selection as well. Although MRA can accommodate the conventional sample selection biases that are often seen in empirical econometrics (Heckman, 1979; Stanley and Jarrell, 1998), it can do much more.

Publication selection, as opposed to sample selection, arises if researchers, editors, or reviewers use statistical significance as one model selection criterion. Publication biases have been identified in the majority of economics areas of research and often have important practical effects (Doucouliagos and Stanley, 2012).[4] Because publication selection is caused by the process of conducting empirical economic research itself, conventional econometrics is incapable of correcting or estimating this effect. Hence, some "macro" perspective is required that looks across an entire research field, and this is precisely what MRA provides. Chapter 4 discusses how MRA can identify, estimate, and correct publication selection bias, and subsequent chapters illustrate and explain how MRA can filter out many other types of bias as well.

The purpose of this book is to introduce the tools of meta-analysis and meta-regression analysis to business and economic researchers unfamiliar with their use. Meta-regression analysis addresses the rising "Tower of Babel" that current economics and business research has become. Evidence-based policy requires a clear and objective assessment of the research record. Without a systematic and objective way to summarize and understand current research, policy discussions will be at the mercy of the subjective interpretation of our empirical knowledge. Moreover, there is a real danger that vested interest or ideology will dominate the discussion and thereby distort policy.

For example, it is clear that both of these forces dominated the anti-regulation atmosphere in the USA that preceded the global 2008 financial meltdown. Alan Greenspan, the former US Federal Reserve chairman, was a disciple of Ayn Rand and a libertarian (Greenspan, 2007; Leonhardt, 2007). Greenspan has been forthcoming about his free-market ideology. A case has been made that it was the opposition to the regulation of derivatives by both Greenspan and the financial industry that led to the worst recession in the USA since the Great Depression (Public Broadcasting Service, 2009), and Greenspan admitted the error of his ideology to the US Congress (Andrews, 2008).[5]

A more positive trend is that governmental agencies are funding dozens of systematic reviews and meta-analyses of their programs and policies.[6] In 2011, the United Kingdom's coalition government has renewed its pledge to protect its international development aid from the spending cuts and to double its international aid commitment. Needless to say, this puts the Cameron government under considerable political pressure, not the least of which comes from their party loyalists (Hennessy, 2011). In a climate of large cuts to domestic programs, it is especially important to ensure that government policies and programs are getting "value for money." Here, too, MRA has an important role to play, because it can offer an objective, comprehensive and rigorous summary and evaluation of what is known, empirically, about a given intervention or policy.

In our view, we are at the dawning of a new era of empiricism in economics and business. Even though the capacity of future empirical methods cannot be fully known, there will remain conflict in what these methods reveal about specific business and economic effects. These phenomena are irreducibly contingent on prevailing cultural and political institutions, and we live in dynamic societies. Thus, an important role for meta-analysis is virtually assured.

If the past is any guide, future systematic reviews and meta-analyses will, on occasion, find that strongly held economic theories are not supported by the weight of empirical evidence. For example, minimum wage raises do not cause lower employment in the US (Doucouliagos and Stanley, 2009) – see Chapters 4 and 5. In other cases, intentionally weak governmental policy (i.e. non-mandatory regulation) will be found to have their intended effects – for example, chief executive pay and corporate performance (Doucouliagos *et al.*, 2012a).[7]

1.2 A historical sketch of meta-regression analysis

One can begin the history of meta-analysis at several points. One choice is the early twentieth-century contributions of the legendary statisticians, Karl Pearson (1904) and R.A. Fisher (1932). Both sought a means to combine separate experiments statistically and rigorously. Because experiments tend to be expensive, the sample sizes employed are often too small to obtain statistically significant results in individual studies. Thus, an obvious statistical solution to economize scarce experimental knowledge is to combine several small-sample experiments to increase their overall statistical power and thereby obtain that all-pervading research goal, "statistical significance."

Pearson's solution was to average the correlation coefficients, while Fisher developed a new statistic that combined p-values. Pearson's approach is very simple and obvious when we look back a hundred years. Nonetheless, weighted averages of correlation coefficients are still used by meta-analysts (see Chapter 3). The elegance of Pearson's solution is that the correlation coefficient is a pure number with no units of measurement, allowing different, but related, outcome measures to be meaningfully compared and combined. This issue of which statistics and measures can be meaningfully combined remains a central issue confronting every meta-analysis. The second advantage of using correlation coefficients is that they reflect the underlying magnitude of the empirical phenomenon in question, not merely its statistical significance (Cohen, 1988).

Fisher's approach is more complex, yet much less useful. It assumes, as the null hypothesis, that all studies have no genuine underlying experimental effect. By doing so, p-values become uniformly distributed and give the Fisher combined probability test:

$$f = -2\sum_{i=1}^{L} \ln P_i \tag{1.1}$$

where L is the number of statistical outcomes or studies in the literature, and P_i is the p-value of the ith the study. This Fisher test is distributed as a chi-squared with $2L$ degrees of freedom under the joint null hypothesis of no effects.

Unfortunately, nearly all applications to economics and business research can produce a significant Fisher test; thus, it has little informative value. Worse, it is likely to be misinterpreted as providing evidence that there is actually some important empirical effect when there is merely excess heterogeneity. Assuming that all individual effects are zero implies that there is no bias or heterogeneity in a given research literature. As discussed above, there are too many misspecification biases in applied econometrics and too much natural variation (or heterogeneity) in economic and business phenomena for the null hypothesis of the Fisher test to ever be true. Furthermore, when the null is rejected, nothing is said about the true magnitude or practical significance of the effect in question. Perhaps, there is simply excess variation among the reported results, and some of this variation is selected? Although some researchers still use this test, we believe that it is fatally flawed for meta-analyses in economics and business (see Chapter 3 for a further discussion).

By the 1970s, some fields of study were already swamped by conflicting findings, and the modern era of meta-analysis was born to make sense of them. Gene Glass (1976) is generally given credit for "meta-analysis" as he introduced this term to contrast his synthesis of all relevant research on a given research question with "primary" and "secondary" statistical analyses:

> Meta-analysis refers to the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies that typify our attempt to make sense of the rapidly expanding research literature.
>
> (Glass, 1976: 3)

Glass was interested in showing that psychotherapy had a beneficial effect. For a couple of decades, the effectiveness of psychotherapy had been in great dispute and both sides resorted to "vote-counting" hundreds of relevant papers (Hunt, 1997). Glass understood that merely counting the number of studies that found a significant treatment effect was not a valid way to accumulate scientific evidence.[8] Rather, Glass offered the "effect size," *g*, as a means to compare the magnitude of the empirical effects reported across studies:

$$g = \frac{\overline{X}_e - \overline{X}_c}{S} \tag{1.2}$$

where the numerator of this ratio is the average difference between the experimental and control groups on some relevant measure of effect, and S is the standard deviation of this measure as seen in the control group. Glass's g is a standardized measure of effect that has no dimensionality; that is, no units of measurement. As such, studies that employ different outcome measures (e.g. different scales of mental health) can be directly combined and compared. Note also that Glass's g preserves the magnitude of this effect, not merely its direction or significance. If the measured mental health of treated patients improves a lot, on average, relative to the background variation in what happens to similar, untreated subjects, g will be correspondingly large.

Now, it is standard practice to report "effect size" in education and psychology as a way to focus on the practical importance of an empirical finding, not only its statistical significance.[9] Cohen (1988) offers widely accepted guidelines for the practical interpretation of effect size. When $.2 < g < .5$, there is a small effect. A medium effect has $.5 < g < .8$, and a large effect is found when g exceeds .8.

Smith and Glass (1977) summarize hundreds of studies of psychotherapy and show that it has a beneficial, if moderate, effect on patients' mental health – on average, $g = .68$. After Glass (1976), the use and further development of meta-analysis slowly blossomed in psychology and medical research, two fields where experiments often give differing results. Meta-analysis is now the accepted method to summarize scientific knowledge; its results are often regarded as "definitive." See Hunt (1997) for a more comprehensive, yet delightfully readable, history of the development and application of meta-analysis.

Empirical econometrics employs a variety of multiple regression techniques to isolate the marginal effect of price, income, some intervention, or other economic variable, holding a myriad of other factors constant. These partial and marginal effects are what typically interests economists. As discussed above, different statistical methods, models, and sets of independent variables are used to estimate the marginal effect in question; thus, we always find much misspecification bias and large heterogeneity among reported econometric estimates. To accommodate and filter out these biases and genuine heterogeneity, Stanley and Jarrell (1989) proposed using essentially the same statistical tools which produce econometric estimates, to summarize and explain the observed variation in these reported estimates. "Meta-regression analysis" was always conceived as a "multivariate" means to summarize and explain multiple regression estimates or transformations of these estimates.[10] Economic meta-analysts believe that observed econometric estimates are the product of complex multifaceted forces, much like the observed economic phenomena, themselves.

Another advantage of MRA is that it uses essentially the same tools and statistical methods as do the econometricians who produce empirical economic estimates. Thus, econometricians have no rational basis upon which to object to the heightened scientific scrutiny that MRAs offer. If there is some fundamental weakness or limitation in MRA, then econometrics will likely suffer from very similar problems.

Economists produce millions of empirical estimates each year, and they are used by policy makers to design critical interventions (e.g. a stimulus package to moderate a recession). Meta-regression analysis takes empirical economics seriously and seeks to improve it. The point of departure of MRA is that empirical economic estimates represent an important phenomenon worthy of deeper examination. Unsurprisingly, economists have been slow to accept this added level of scrutiny. After all, what producer embraces an objective and critical assessment of his products? Nonetheless, MRA has been widely accepted in recent years just as quality assurance is conventional practice in manufacturing.

To provide a rough sketch of the trajectory of the discipline's adoption of meta-analysis, we searched EconLit for "meta-analysis" or "meta-regression" in either

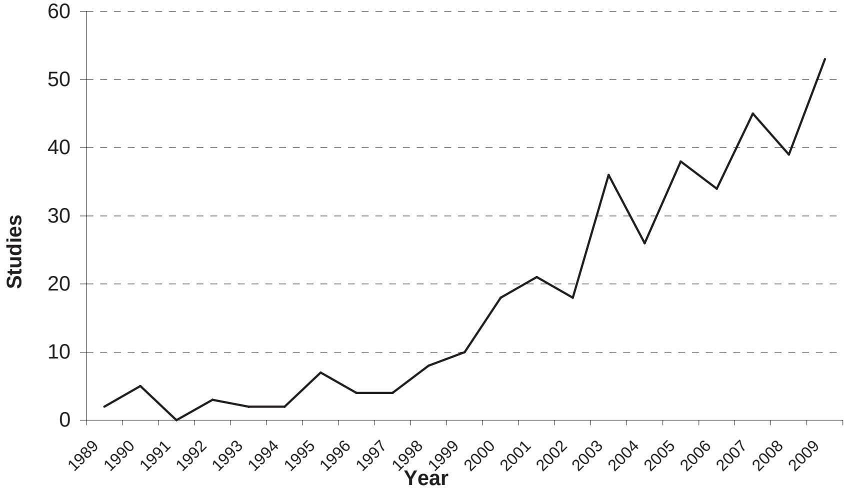


Figure 1.1 Meta-analysis in economics over time

the abstract or title among all papers that also concern "economics." Figure 1.1 plots the growth of published meta-analyses in economics over its first 20 years, while Figure 1.2 shows that a simple exponential growth model provides a rather good fit ($R^2 = 0.88$). The adoption of meta-analysis in economics is, on average, growing at 18 percent per year. We know that these numbers of meta-analyses of

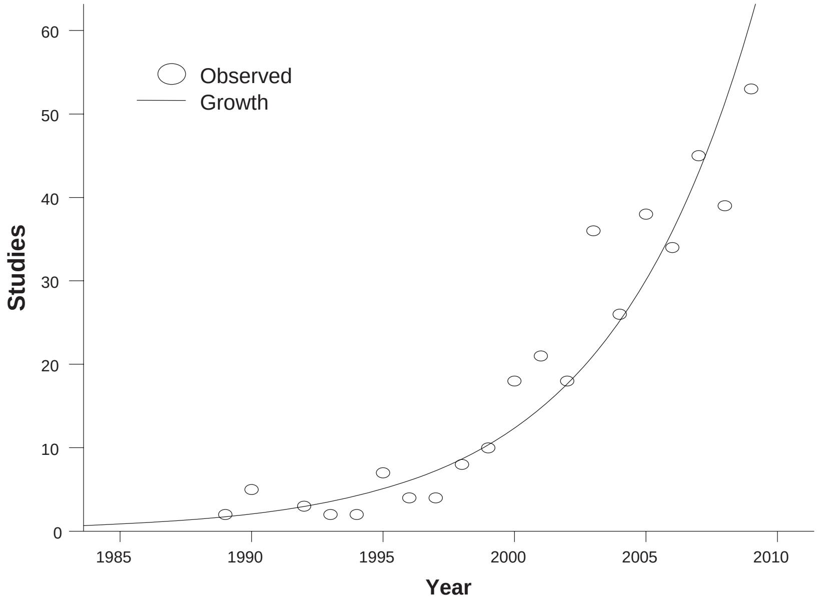


Figure 1.2 The exponential growth of meta-analysis in economics

economics research represent a considerable underestimate because several of our papers are not included and the same search of Business Source Premier uncovers four times as many meta-analyses.[11] In any case, there have been hundreds of (perhaps as many as a thousand) meta-analyses conducted on empirical economics, and there is sufficiently strong momentum and growing policy interest for this trend to continue for the foreseeable future.

This book offers the first comprehensive guide to conducting and understanding the type of meta-analysis (meta-regression analysis) especially designed for econometric studies. Although there are a number of books on meta-analysis, they all concern RCTs or similar types of research that are fundamentally different than applied econometrics. Interest in econometric meta-analysis has sufficiently matured to merit its own guide.

1.3 Practical examples

In order to ensure that this book remains practical and realistic, we frequently illustrate these methods with examples of actual meta-analyses. In particular, four published meta-analyses will be used consistently throughout the following chapters to illustrate the issues, methods, and statistical analyses involved in the meta-analysis of economics and business research. These are: the effects of unions on productivity (Doucouliagos and Laroche, 2003), residential water price elasticities (Dalhuisen *et al*., 2003), the value of a statistical life (Bellavance *et al*., 2009), and minimum wage elasticities (Doucouliagos and Stanley, 2009). These four areas are selected for a variety of reasons. First, we must have access to the full set research data employed; otherwise, a comprehensive range of meta-analytical statistics could not be computed. Second, we wish to display a wide range of meta-analyses from different areas of research. For example, union productivity was selected because it is one of the best examples of the absence of the distorting influence of publication selection, and we have access to the data. The other three were thought to be especially important for their policy implications.

The magnitude of residential water price elasticities is critical to a city manager or environmental planner who wishes to use price or taxes to conserve water. This is a policy that is already important in many areas of the world and is likely to become even more crucial in the not too distant future. The larger the magnitude of this elasticity, the more effective price and/or tax rises will be in conserving water. Unfortunately, as we demonstrate in Chapter 4, we find that water consumption is quite insensitive to rises in prices and taxes once likely publication biases are accommodated. Thus, such conservation policies are not likely to be as effective as a conventional yet comprehensive reading of this research literature would lead you to believe.

The value of a statistical life has even wider policy implications on almost all health and safety policies, regulations, and projects. Regardless of one's subjective views about the value of a human life, practical choices must be made concerning which health and safety laws and regulations to adopt and in which health and safety projects to invest. In these necessary calculations, the value of life is often the most critical single parameter. For example, the acceptability of environmental regulations is often decided by the number of lives saved and the value of those lives. Rather than using some arbitrary or politically chosen value, the value of a statistical life (VSL) is indirectly estimated by observing how workers and citizens voluntarily accept higher risks, buy insurance, or reveal their preferences on surveys.[12] Bellavance *et al*. (2009) meta-analyze only one general source of such VSL estimates, hedonic wage equations. A hedonic wage equation estimates the risk–wage tradeoffs that workers make, and from these a VSL may be imputed (Viscusi, 1993). Here, too, we find that selection bias has a large practical effect on this important magnitude, inflating it by a factor of 5 (see Chapter 4).

Lastly, we use the employment effect of minimum wage increases as a reappearing example, largely because its meta-dataset is so rich. We have collected 1,474 comparable minimum-wage elasticities all for the USA and have coded a couple of dozen potentially relevant research dimensions to help explain the wide range of employment effects reported in the research literature (Doucouliagos and Stanley, 2009). Needless to say, whether or not minimum wage increases have an adverse effect on employment also has clear policy implications. It is this adverse employment effect that is used by opponents to block increases in the US minimum wage as it comes up for a vote every few years. And minimum wage laws are found across the world. Our MRA goes against conventional wisdom and most of the reported research studies. We find robust evidence for the absence of any practically significant adverse employment effect from minimum wages in the USA (Doucouliagos and Stanley, 2009) – see Chapters 4 and 5.

Needless to say, many other important research dimensions are found in these four MRAs and discussed in detail below. Although these four areas of research will be used throughout the book, we supplement them with many other tangible meta-analyses to illustrate particular issues and statistical methods when they are more revealing or more germane. Hundreds of meta-analyses have been conducted in economics, and what is seen in these four meta-analyses is broadly characteristic of this larger research literature.

1.4 Plan of the book

Meta-regression analysis is best seen from a broad perspective. It offers a framework that can *simultaneously* be used to: summarize and qualifying estimates of policy-relevant parameters; correct these estimates for any number of potential biases inherent in observational economics research; test economic theories; explain heterogeneity; model the research process itself; and give direction to future empirical investigation. The following chapters illustrate how MRA can achieve each of these objectives.

Chapter 2 offers strategies for identifying and coding empirical economics and business research. Conducting these activities in a way that is replicable by others is crucial for the quality and scientific status of the resulting meta-analysis, and they represent 90 percent or more of the effort involved.

Chapter 3 discusses simple descriptive statistics and graphs that have been found useful in summarizing research. Its purpose is to paint a clear, if coarse-grained, picture of what empirical inquiry has thus far uncovered and thereby help generate hypotheses that can be more rigorously tested and investigated by the full panoply of statistical techniques.

Chapter 4 introduces meta-regression methods that identify and correct publication selection bias. Most economics research exhibits “substantial” or “severe” publication selection bias (Doucouliagos and Stanley, 2012). Yet, conventional econometrics, no matter how rigorous or comprehensively applied, can be overwhelmed by this bias and is powerless to correct it.

Chapter 5 shows how multiple MRA is often employed to explain economic research and its excess heterogeneity. Like economic phenomena, economics research cannot be accurately summarized or fully understood without explicitly accounting for a multiplicity of complicating factors.

Chapter 6 offers a theory of meta-regression analysis and a rigorous demonstration that study quality need not affect the findings of a MRA. It more deeply explores MRA models for within-study dependence and publication selection.

Chapter 7 further describes alternative objectives for performing systematic reviews and how they shape the way MRAs are conducted or applied. It also considers additional complexities to the structure of empirical research and how to model them statistically.

Chapter 8 concludes and summarizes the book.

2 Identifying and coding meta-analysis data

> The commonly held belief that research progress will be made if only we “let the data speak” is sadly erroneous. ... it would be more accurate to say that the data come to us encrypted, and to understand their meaning we must first break the code.
>
> (Hunter and Schmidt, 2004: xxxi)

Empirical studies and their estimates are scattered throughout a complex research landscape. Some are published in prestigious peer-reviewed journals; others exist only in unpublished working papers, dissertations or online. Still others are known only to the researcher who produces them, never to be seen by any other scholar. Relevant empirical estimates are generated from sophisticated structural econometric models, reduced-form models, and also simple bivariate comparisons. Studies differ widely in terms of the control variables, data sources, and estimation techniques employed. This multidimensional nature of research makes deriving clear inferences and policy advice difficult. What we need is an efficient set of tools to summarize, integrate, correct, and evaluate research findings.

Enter meta-analysis. Like all statistical techniques, data fuels meta-analysis. However, “data” in the meta-analysis context are the complex products of the research process. Typically, meta-data will be comprised of estimates of some economic association (also known as “effect sizes”) *linked* to key dimensions of the research process that produced these effects. Like any good empirical study, meta-analysis commences with a theory, or a group of theories, that predict associations. These theories are next investigated empirically in a research literature. While it is this empirical literature that meta-analysts explore, it is the underlying economic theory that shapes this empirical inquiry that is of ultimate interest to economists. Hence, an indispensable component of a meta-analysis is a thorough understanding of the underlying economic theories. This understanding will shape the meta-analyst’s search for studies, coding of research, subsequent statistical analysis, and ultimately the interpretation of the meta-analytic statistics produced. Meta-analysis begins neither with statistics nor data, but rather with a clear understanding of economic ideas.[1] Theory provides the topographical map of the terrain to be explored, while meta-analysis provides the tools for extracting the precious ores, should they be present.

This chapter focuses on the collection of the data that defines meta-analysis. In particular, we discuss where to collect the data and what information to collect. Chapter 3 explores alternative ways of summarizing research findings. More complex statistical meta-analyses will be explored in subsequent chapters.

2.1 Identifying studies

Identifying studies to include in the meta-analysis means conducting a literature search to identify empirical studies that offer estimates that are comparable both within and between studies.[2] A critical feature of this search is that it should be as comprehensive as possible. It is very important that the meta-analyst herself introduces no systematic bias into the data (Stanley, 2005a). In our view, the central task of meta-regression analysis (MRA) is to filter out systematic biases, largely due to misspecification and selection, already contained in economics research. Thus, to systematically select the research literature runs the risk of defeating the very purpose of meta-analysis in economics. We return to this important topic below and in subsequent chapters.

Most researchers are familiar with conducting searches through traditional qualitative literature reviews. Such reviews typically cover only a fraction of the available studies, and often researchers merely re-review the studies that are most commonly known. Meta-analysis, however, requires additional search efforts. Meta-analysis should be conducted on the *population* of studies that satisfy a set of search criteria, or at least a representative sample of them. Most meta-analyses in economics we have reviewed have been conducted on a population of studies, or as near the population as feasible.[3] However, in their review of 140 meta-analyses in environmental economics, Nelson and Kennedy (2009) found that most did not involve the population of studies, and the gray literature was underrepresented.

Identifying the population of studies is not a trivial matter. Where a literature is known to be enormous, tighter exclusion criteria may reasonably be adopted. For example, the empirical growth literature contains thousands of studies. When conducting a meta-analysis of this literature, it might be more practical and cost-effective to restrict the search to only those studies published after some year, say 2000.[4] Other restrictions are also possible. For example, the search can be restricted to studies examining one specific growth effect, using panel data, or to those modeling endogeneity. Another option is to take a random sample. This approach is rare in economics, where the preference has been to code the population of studies. Taking a random sample of studies might be worth considering where there exist a very large number of studies. See Abreu *et al*. (2005) for an application.

Existing narrative literature reviews offer a nice base from which to begin the search for studies. There are numerous academic search engines. Searches in economics typically commence with EconLit, supplemented with search engines such as Google Scholar and Scopus or similar search services. Knowledge of the underlying theory is indispensable to identify keywords for the search. As an example, consider a search for the effects of economic growth on attracting foreign direct investment (FDI). It will be insufficient to simply use the words “growth” and “FDI”. It will be necessary to understand all the determinants of FDI and include these too in the search. For example, a whole vein of studies explores the effects of taxation on FDI, and includes growth as a control factor. Limiting the search to “growth” and “FDI” may not reveal such veins of research, potentially resulting in systematic bias.

Many relevant studies are detected by a careful reading of the primary studies themselves,[5] especially literature review sections of these studies and their reference lists. Such careful reading often reveals studies that are missed by search engines, usually because the papers’ title or abstract does not contain the keyword that was searched. Citation is another useful avenue to find additional relevant research studies. It is always a good idea to check the studies that have been cited by an identified study, as well as the references found in these citations. Citation searches can be conducted through search engines such as: Google Scholar, the Social Science Citation Index, Scopus, or through *Publish or Perish*.[6]

Taking the time to detail the search strategy employed (e.g. as an appendix to the paper) is important for independent validation. Meta-analyses in medicine report the exact keywords and databases that were searched. More recent meta-analyses in economics appear to be adopting this recommended practice.

2.1.1 Selection criteria for studies included in meta-analysis

It is important that the meta-analyst adopts an explicit set of selection criteria. These criteria define the population of studies that will be collected and analyzed. These criteria should be stated explicitly and clearly in the published meta-analysis. The studies included in the meta-analysis should be so similar that their differences can be coded. Obviously, if more than one hypothesis is to be tested, then a separate search might need to be undertaken for each hypothesis. Stroup *et al*. (2000) provide a useful checklist for the search strategy in particular, and meta-analysis more broadly. Most of these are relevant also to economics. Several systematic review and meta-analysis organizations (the Cochrane and Campbell Collaborations and MAER Network),[7] have established their own guidelines or will likely do so in the near future. Higgins and Green (2008) give excellent guidance for systematic reviews of medical research.

All meta-analyses should begin with an initial search for *empirical* studies: studies that do not report an estimate cannot be statistically analyzed.[8] Typically, this will mean identifying applied econometric studies. That is, most of the time, the meta-analyst of an empirical economic literature will search for regression-based estimates of an effect. At the minimum, this will mean collecting studies that report regression coefficients, sample size, standard errors and/or *t*-statistics. This information enables the most basic statistical meta-analysis.

In some cases, however, it might be necessary to include non-regression empirical studies. For example, an important area in leisure research is user satisfaction with outdoor recreational facilities (Shinew *et al*., 2004). While some of the empirical studies in this field report coefficients from logistic regressions, many do not. An alternative approach is to use reported sample proportions. The focus of the meta-analysis is then the proportion of satisfied recreational users rather than estimates of some marginal effect estimated from econometric analysis.

It is essential that the collection of studies and the coding of findings cover the same empirical relationship. In most cases, it will be necessary to add additional exclusion criteria to ensure the comparability of empirical estimates. Typically, only a fraction of the studies identified by this initial search can be coded in a compatible manner and thereby included in the meta-analysis. In general, it is the completeness and comparability of the reported research, rather than the meta-analyst herself, that determines which findings will ultimately be comparable and thereby able to be meta-analyzed. As an additional check, the meta-analyst can list the criteria that need be satisfied by included studies. Studies could then be ranked on the basis of satisfying these criteria, and the robustness of the MRA can be assessed in the face of increasingly more stringent subsets of comparability.

Incomplete reporting of research

It is critical that the way in which the research process was conducted is understood, communicated, and coded. It is not uncommon for some studies to report effect sizes but fail to provide enough information on the type of data used, on the construction of key variables, or on the way in which critical modeling issues were handled. These studies need not be omitted by the meta-analyst, but they will drop themselves out in more complex MRA when the associated moderator variables have missing values.[9] Most critical is that the dependent and key explanatory variables are described fully and that measures of the effects are comparable. Most primary empirical studies will adopt an established and reliable measure of the variables of interest. However, some authors might construct their own measures, thus calling into question whether these studies can be included in the meta-analysis. At a minimum, such differences in measures need to be explicitly coded, and their effect explored through MRA (see Chapter 5).

Non-English studies

Most authors will include only studies written in English. In other disciplines, most notably in medicine, some effort is taken to include non-English studies. We do not, in general, see this as a critical issue in economics. Most empirical economics papers are actually written in English, so that any bias resulting from omitting non-English studies should be of a second order. Moreover, it will be a rare case where only the effect sizes and standard errors need to be identified in a non-English study. Simply getting an English translation of the reported estimates will be insufficient. It is critical that studies are understood clearly.

There are, of course, obvious exceptions. For example, if the aim was to assess the effects of economic policy in say Latin America, it might be necessary to access non-English papers. Likewise, if the study's aim is to analyze labor demand elasticities in France, then studies published in French would need to be included.

Obscure studies

The same principle applies to hard-to-get studies. Some studies are published through obscure journals or working paper series that are very difficult to access. The inability to obtain and hence include these studies is unlikely to affect the findings of a meta-analysis. Indeed, it is our experience that the central findings of meta-analyses are remarkably robust to marginal changes to the definitions of the population of studies, the data, or the coded moderator variables. However, influential estimates, in a statistical sense, do occasionally exist. When influential estimates are found in the "gray literature" their inclusion is suspect. Even when contained in our best academic journals, the influence of any one estimate should be minimized by employing robust statistical techniques. These recommendations may seem strange to medical researchers, where one large randomized clinical trial may be more reliable than all of the remaining research combined. In economics, even the most rigorous and sophisticated research study may have omitted a critical variable or somehow misspecified the estimation model and thereby report biased estimates due to unavoidable data or methodological limitations.

Binary dependent variables

Effect sizes need to be comparable. In general, it is not possible to combine estimates from binary regressions (e.g. probit and logit studies) with estimates from continuous variable studies (e.g. ordinary least-squares studies). Hence, these studies are excluded from a meta-analysis of a continuous effect (which in most cases is the larger group). However, where there are enough studies, a separate meta-analysis of the binary regression results using the log-odds ratio can be undertaken.

An alternative strategy is to change the focus of the meta-analysis away from an analysis of the effect size, to whether a certain type of result is found (e.g. whether a statistically significant effect was reported). This then enables the meta-analyst to combine all studies together using meta-probit analysis. Examples of this approach include public subsidies and business research and development (García-Quevedo, 2004) and the evaluation of active labor market policies (Card *et al*., 2010).[10] It is our view that a meta-logit/probit should be undertaken with caution. Taking a continuous variable and arbitrarily dichotomizing it (e.g. significant or not significant) is likely to introduce a spurious structure into the data that does not correspond to any underlying reality. There is the danger that the significant moderator variables identified by a meta-logit analysis will reflect mere correlation with the publication selection process rather than any genuine characteristic of the underlying economic phenomenon studied. See Chapters 4 and 5 for a discussion of publication selection bias and especially the use of "*K*-variables" in multiple MRA (Chapter 5).

In some cases, there might be insufficient estimates from either binary or continuous regressions, but there might be many studies that report whether a certain result was found. A meta-logit/probit might then be applied to explore the study characteristics that lead to certain results. For example, most of the studies that have looked at the issue of collective action for natural resource management have been case studies (see Poteete and Ostrom, 2008). Estimating a meta-logit for this literature makes perfect sense because it uses, and preserves, all the available information, and is less likely to produce artificial patterns where there were none. A meta-analysis of such literatures may still be somewhat problematic because the data, or the information, is so rough to begin with. But that is the nature of qualitative (or case study) research.

Same estimates

Some studies have the same author(s), use the same data and report the *same* estimates as previously published in an earlier paper. These should not be included.[11] Some studies are pure replications: they use the same data, the same estimates, but are produced by different authors. Some meta-analysts choose to exclude these, while others include them. If replication studies are included, their replication status needs to be coded.

2.1.2 Unpublished papers

An important consideration is the treatment of unpublished papers and reports. This is sometimes called the "gray literature" (Loomis and White, 1996). Many meta-analyses have been conducted on published studies only. The main reason given for this is that published studies have gone through the refereeing process and, hence, should be of greater quality than unpublished papers. This is part of the methodological quality issue, that is, the argument that meta-analysis should include only those studies and estimates that are of high methodological quality. The inclusion of low-quality estimates might taint the meta-analysis. We return to this issue in Section 2.5 below.

As we see it, a downside of not including unpublished studies is that they tend to be newer studies, using newer data and newer estimators. They thus might capture structural changes in the effect or fresh thinking on how to best model the effect. In such cases, ignoring unpublished studies might result in biased and inferior estimates. In a large and mature literature that is free of publication selection bias, the risks of omitting unpublished papers are probably minimal. However, in a small, rapidly emerging field, it is prudent to consider unpublished estimates.

Furthermore, all studies, regardless of their quality, are helpful, and sometimes essential, in the statistical identification of specific research dimensions that are responsible for the wide variation found among the reported research results (Stanley *et al*., 2008). Routinely, MRA uses independent variables that are indicators of the observed differences in research methods, models and data, and hence quality. To remove unpublished papers systematically that tend to be of lower quality or less rigor might in some cases render MRA incapable of understanding the observed variation of research results because there will be insufficient variation in the independent variables that represent differences in research methods, models and data. Basic econometrics recognizes that we always want the largest possible variation in the explanatory variables to obtain reliable regression statistics. For example, if we were to remove unpublished papers from the meta-analysis of the efficiency-wage hypothesis, we would not be able to identify the importance of including (or omitting) a measure of the capital stock in the production function. Yet, this omission is revealed to have a practically large effect on the reported magnitudes of the efficiency-wage effects when both published and unpublished studies are included (Krassoi-Peach and Stanley, 2009).

Another downside to the unnecessary restriction of meta-data concerns the issue of publication bias, which is the main topic of Chapter 4. Several authors have justified the inclusion of unpublished papers on the basis that this will resolve the issue of publication bias (e.g. Zelmer, 2003; Égert and Halpern, 2006). However, it is not sufficient to merely include unpublished papers. It is typically necessary to test publication selection bias formally and correct a literature for potential publication bias, whether or not unpublished studies are included.[12]

In our experience, there is publication selection bias even among unpublished papers and no detectable difference in quality between published and unpublished papers as measured by the objective statistical criterion of precision. In more cases than one might expect, the published research literature will contain too few comparable estimates to conduct the needed multiple MRA. Examples include efficiency wages (Krassoi-Peach and Stanley, 2009) and the effects of advertising on the onset of alcohol consumption (Nelson, 2011). Thus, unpublished papers should be routinely included in a meta-analysis. The meta-analyst can always code for the publication status of the study and see whether the exclusion of unpublished papers practically affects the meta-analysis results. When they do, this is evidence for the inclusion of unpublished studies unless the meta-analyst can make a strong case, on objective grounds, that the unpublished studies are materially of lower quality.

Unpublished studies come in several varieties. At one end, there are studies that are published in recognized and highly respected series, such as the NBER working papers. After this come unpublished doctoral dissertations and numerous departmental working paper series. At the lower end are unpublished manuscripts presented as non-refereed conference papers.[13] Depending on the research area, studies by consulting firms and government agencies might also be considered. Some of these are highly reputable (e.g. those from a national central bank) while others might be mediocre. Hence, if unpublished studies are to be included, the meta-analyst has to form some judgment as to which unpublished studies to include. Again, these differences can all be coded, and their potential effects on the meta-analysis can be objectively assessed through MRA (the topic of subsequent chapters).

One danger with the use of unpublished studies is that there is a risk that the estimates in the published version might change. If the refereeing and review process works well, these changes should improve precision and accuracy. Yet, there is a paradox here. In a large, mature and well-established literature, such changes are unlikely to affect inferences from meta-analysis, and the exclusion of unpublished studies is unlikely to affect the results. In contrast, in a small and emerging literature, the exclusion of unpublished studies might affect inference. However, this also allows for the possibility that inference will be affected if the results in the unpublished studies change in the published version of the study.[14]

2.1.3 Published papers

There is also the issue of which published papers to include. Academic journals are deemed to be the warehouses and guardians of scientific knowledge. Hence, it comes as no surprise that all meta-analyses conducted in economics so far have relied heavily on academic journals. Estimates, however, can also be published in research books, chapters of edited books, published doctoral dissertations, published government reports and even professional journals. Hence, the meta-analyst needs to decide if the search will be restricted to only refereed academic journal papers, or whether it will be broadened to other publication outlets. Both approaches have been adopted in the literature. Our advice, in all cases, is to err on the side of inclusion. Differences suspected to be important can always be coded and explicitly included in any MRA.

Does it make a practical difference?

In our experience, if a comprehensive search is conducted and the references cited in the literature are included, then systematic bias in the overall corrected effect from not including other less well-known and uncited studies is minimal. In contrast, differences in the types of studies included can make a substantial difference in which research dimensions are found to make a significant contribution to our understanding of the variation among reported research results. Therefore, we believe that the extra effort of searching for unpublished studies is worth the search and coding costs. First, the meta-analyst can take some comfort that she has a comprehensive dataset of the literature.[15] Second, when exploring bias in a literature, particularly publication selection bias, it is important to employ a full dataset.[16] Third, the meta-analyst will have greater information and degrees of freedom with which to explore the heterogeneity between estimates.

A priori, it is not obvious whether the inclusion of unpublished studies will make a difference. If unpublished studies have been collected, it is probably wise to undertake a sensitivity analysis of the meta-analysis, that is, conduct the meta-analysis with and without unpublished studies.[17] Many studies find no difference in the results between published and unpublished papers. Examples include the studies by Zelmer (2003), Fidrmuc and Korhonen (2006), and Klomp and de Haan (2010). In contrast, Doucouliagos and Stanley (2009) find that published studies report larger minimum wage effects, while Alston *et al*. (2000) find that journal papers report lower rates of return to agricultural research and development.

Kluve and Schaffner (2008) find that papers published in journals report smaller values of a statistical life in most of their meta-regressions. Note that if the meta-analyst adopts some sort of a weighting scheme using journal quality, then the unpublished studies will, by definition, be given a zero weight (see Chapter 3).

In order to exclude a study from a meta-analysis it must first be identified and objective criteria established to justify the exclusion. Excluding a study from a dataset means assigning a zero weight to it. Meta-analysts make a serious effort to consider all empirical studies on a given topic. Thus, there must be good objective reasons to exclude any class of studies.

2.1.4 How many studies to collect?

We aim for a comprehensive assessment of a literature for its own sake. The object of the inquiry may be economics research itself (Stanley *et al*., 2008). If so, all relevant research should be coded and included explicitly in the meta-analysis. Hence, great effort needs to be taken to identify all studies that are relevant. Doucouliagos and Stanley (2012) found that the average number of studies included in the 87 meta-analyses they reviewed was 41, with the median being 35. In some literatures, however, there will be literally hundreds of studies that need to be collected. For example, in their extensive meta-analysis of gender wage differentials, Weichselbaumer and Winter-Ebmer (2005) collected estimates from 236 studies, while Gallet (2010) collected 3,357 estimates from 393 studies.

2.2 What data to collect

The collection of data can occur at three levels. At the bare minimum, a meta-analysis requires that data be collected on the association that is of central interest and some variable that will be used to weigh the associated effect sizes, *essential*. The effect size will be some measure that quantifies the association. In Section 2.3, we list the common effect sizes used in economics. The second set of information relates to details about the study and the research process, *typical*. The third set involves the collection of study-invariant information, *value-added*.

Essential

The most basic meta-analysis will involve a simple weighted average and/or a simple linear regression model. This requires that data be collected on effect sizes, their sample sizes and standard errors. It will also be necessary to code the name(s) of the author(s), the title of the paper and publication outlet.[18]

Typical

Most meta-analysts collect data on the types of data used in the study, the estimation technique used and the econometric model structure. Examples of data differences include: cross-sectional, panel, or single country and whether establishment, firm or industry level data were used. Further data differences involve: the country under investigation, the type of industry (e.g. manufacturing or services), and the time period under investigation.[19] Experience has shown that the functional form (e.g. double log) and the exact model specification used (the control variables included) are important dimensions of the original research to code and to model. The estimator used can also be important. Does the study use ordinary least squares? Does it control for endogeneity? If panel data are used, does the study control for fixed effects? Where there are rival theories regarding an effect size, it might be useful to include information on the theory tested.[20] It is also important to note omissions of relevant independent variables, which can bias the original research findings, and the causal or error structures (e.g. endogeneity) modeled in each primary study. Estimating and correcting empirical economics for such potential biases was, in fact, the original intent of MRA (Stanley and Jarrell, 1989). This topic is discussed in further detail in Chapters 4 and 5.

Meta-regression analysis has the potential to correct the original econometric research for a variety of misspecification, omitted-variable, and other biases. Thus, it is very important to code studies and estimates for obvious omitted relevant variables that might bias the original estimates. For example, omitting a measure of capital in a production function is an obvious omission that might seriously bias the remaining estimates. For the efficiency-wage effect on productivity, omitting a measure of capital reduces the corrected estimate of the efficiency-wage effect by nearly half of its typical value, making a practical and statistically significant difference (Krassoi-Peach and Stanley, 2009).

Value-added

One of the great advantages of meta-analysis over both the original econometric research and conventional narrative reviews is that it can add new and relevant information unavailable to the original study to explain variation in research findings. As mentioned above, MRA can be used to control and correct for omitted-variable bias (see Chapter 4). But these omissions may have been unavoidable in the original study due to data limitations. In many economics databases, information on known relevant variables simply does not exist. Often study results will be influenced by factors that are "study-invariant." That is, factors that are constant for a given study but vary across studies. Only the meta-analyst can model and estimate the effects of these factors on econometric findings because a given "study-invariant" dimension, by definition, does not vary across the data within a given study.

An example of how study-invariant data may be added is found in Jarrell and Stanley (1990), where the unemployment rate is added to the MRA that explains the union-wage premium. Due to the largely cross-sectional nature of early union-wage studies (Lewis, 1986), the unemployment rate or any other cyclical indicator does not vary in most of the data used by this research literature. However, from the perspective of a meta-analysis, the unemployment rate is easily added and found to explain successfully some of the difference among reported union-wage premiums (Jarrell and Stanley, 1990). A second example adds study-invariant spatial data to the meta-analysis of the willingness to pay for land preservation (Johnston and Duke, 2009). This information is unavailable in the original studies because they involve surveys of willingness to pay for preservation of a specific geographic site. Johnston and Duke (2009) find that adding this study-invariant geographical data back into their MRA has practically important effects on the estimates of willingness to pay. Furthermore, adding a geographical dimension is especially important when meta-regression estimates are used for benefit transfer to unstudied sites at new geographical locations (meta-analysis for benefit transfer is discussed in Chapter 7).

It has become routine for meta-analysts to include the average year of the data and/or the year that a study was published, and other study-invariant dimensions, as a means to account for potential trends or path dependencies in research.[21] Increasingly, it is becoming common to collect information on citations received and journal impact factors.[22] These are not collected from the study itself and are again study-invariant. Further study-invariant measures may include more “socio-economic” factors such as: the study’s authors, their gender (Stanley and Jarrell, 1998), their funding sources (Doucouliagos and Paldam, 2010), and their links with other researchers (Doucouliagos and Laroche, 2003).[23] Meta-regression analysis can be used to study the socio-economic process of economics research itself (Stanley *et al*., 2008). The potential for meta-analysis to add value to existing research is nearly limitless.

In the process of analyzing data, new directions and dimensions sometimes emerge that the analyst would like to explore further and more objectively. This data might not have been coded initially. Thus, it is not uncommon for the meta-analysts to go back to the original studies. Furthermore, referees might require added dimensions to be explored to ensure that the central findings of the meta-analysis are robust, and this too might require additional data collection. So be prepared to revisit your coding of the original research.

We provide concrete illustrations of the information needed to be coded when we look at multiple meta-regression models in Chapter 5. All this coding, however, can impose a heavy burden on degrees of freedom. We return to this challenge in Chapter 7.

2.3 Effect sizes in economics and their standard errors

In the meta-analysis of areas like medicine, a wide range of effect sizes is used including: Cohen’s *d*, the odds ratio, Glass’s *g*, log-odds, and log-risk ratios. These are rarely used in economics and, hence, are not discussed further here. The interested reader should consult anyone of a number of standard references such as Sutton *et al*. (2000), Lipsey and Wilson (2001), Whitehead (2002), Hunter and Schmidt (2004), and Borenstein *et al*. (2009).

What should the effect size measure? Ideally, we want a measure of the *economic* effect of a particular variable thought to be conditionally invariant. There is an important difference between *statistical* effects and *economic* effects.

Statistical effects, such as zero-order correlation coefficients and partial correlation coefficients, are unitless measures of an association between two variables. An economic effect, on the other hand, measures the main effect of economic interest. These are typically elasticities, or some other measure that captures the *percentage change* in the dependent variable or some measure of the marginal effect.

The most common approach to meta-analysis in economics is to extract effect sizes from reported econometric models (see Stanley and Jarrell, 1989). A typical econometric model (say, using panel data) takes the form of linear regression:

Linear functional form:

$$Y_{it} = \alpha_0 + \alpha_1 X_{it} + \sum_j \alpha_j Z_{jit} + u_{it} \tag{2.1}$$

where Y is the dependent variable, X is the explanatory variable whose association with Y is the main economic effect in question, Z is a vector of other variables that might affect the dependent variable, u is the random error term and i and t index the cross-section and time period, respectively.

Another common functional form is the log-log:

Log-log functional form:

$$\ln Y_{it} = \alpha_0 + \alpha_1 \ln X_{it} + \sum_j \alpha_j \ln Z_{jit} + u_{it} \tag{2.2}$$

Log-linear functional forms are also frequently found in economics research:

Log-linear functional form:

$$\ln Y_{it} = \alpha_0 + \alpha_1 X_{it} + \sum_j \alpha_j Z_{jit} + u_{it} \tag{2.3}$$

Interest, in all cases, lies in either the estimates of α_1 or on the marginal effect of X on Y, which is a function of α_1. The search reveals the group of studies that report estimates of α_1 or the marginal effect of X on Y. Effect sizes in economics are typically computed from regression coefficients. For these to be included in the meta-analysis, they need to possess two important properties. First, the effect should be a partial one; that is, it should measure the effect of one variable on another, holding other factors constant (the familiar *ceteris paribus* assumption of economics).[24] Second, the effect should be comparable within and between different studies (Stanley and Jarrell, 1989; Becker and Wu, 2007).

2.3.1 Regression coefficients

The fundamental requirement that effect sizes be comparable across estimates will usually rule out the direct use of regression coefficients, unless the scale and measures are identical.[25] Exceptions include: where all studies use the same scale (e.g. estimates of the marginal propensity to consume), elasticities from double-log (log-log) econometric models, and in some cases semi-elasticities where log-linear econometric models are standard – for example, the effect of currency unions on trade (Rose and Stanley, 2005).

2.3.2 Zero-order correlations

The zero-order correlation coefficient (or the simple correlation) is widely used outside economics (see Hunter and Schmidt, 2004). This is a measure of the degree and direction of the association between two variables. It is the most widely used effect size in management research (e.g. Tosi *et al*., 2000) and is used occasionally in marketing (Brown and Stayman, 1992). It is not often used in economics because it does not capture the main association of interest to economists – the marginal effect. Indeed, it is possible for the simple correlation to give an entirely false picture of the underlying association. It is not uncommon to find a positive simple correlation when the actual association is inverse. As a consequence, simple correlations are not widely reported in empirical economics research.

For example, in the context of a demand function, a simple correlation between the own price and sales of a commodity could be positive, due to inflation or to aggregate income growth, whereas the actual conditional association is inverse. This arises because a simple correlation does not control for important effects such as income, advertising and the prices of rival goods. This common economic problem of aggregation is also well known among statisticians and meta-analysis as "Simpson's paradox." Simpson's paradox materializes when two conditional correlations of some relation are positive (or negative) but reverse signs when the unconditional correlation is considered (Pavlides and Perlman, 2009). This is, in part, why econometricians focus sharply on conditional effects. Meta-regression analysis goes a long towards removing Simpson's paradox by including moderator variables reflecting whether an important dimension has been omitted in the original econometric analysis and by coding and adding study-invariant variables to the MRA that the original study could not investigate.

Hence, if the simple correlation were chosen as the effect size, only a fraction of the empirical studies can be included in the meta-analysis. Consequently, there is the real risk of a biased meta-sample. The studies that choose to report the zero-order correlation might not be representative of all the studies that have been conducted on the economic effect in question.

Nonetheless, there are some fields where simple correlations are reported and where meta-analysis can be conducted. Examples include Fidrmuc and Korhonen's (2006) meta-analysis of business cycle correlations in central and eastern European countries and Tosi *et al*.'s (2000) meta-analysis of the chief executive pay–performance association.

2.3.3 Partial correlations

The partial correlation coefficient is also a measure of the strength and direction of the association between two variables, but it holds other variables constant.

That is, it provides a measure of association, *ceteris paribus*.[26] Partial correlations are rarely reported directly in primary economic studies. Hence, they need to be calculated from the conventionally reported regression statistics.

The calculation of the partial correlation coefficient, r, is straightforward:

$$r = \frac{t}{\sqrt{t^2 + df}} \tag{2.4}$$

where t denotes the t-statistic of the appropriate multiple regression coefficient, and df reports the degrees of freedom of this t-statistic.[27] Sometimes, when there is a negative effect, the t-statistic is imprecisely reported without its minus sign in the primary literature; thus careful reading is essential.[28] The standard error of the partial correlation is given by $\sqrt{(1-r^2)/df}$.

The statistical significance of partial correlation coefficients can be tested by using the same t-statistic used for its associated regression coefficient. However, we are rarely interested in the statistical significance of an *individual* reported effect, whether measured by a regression coefficient or by r. Rather, meta-analysis seeks to identify patterns across studies and the underlying message of our *accumulated* scientific knowledge.

The key advantage of using the partial correlation coefficient is that it is a unitless measure, allowing the partial correlations from one field or study to be readily compared to partial correlations in some other study.[29] Secondly, partial correlations can be calculated for a larger set of estimates and studies than almost any other effect size measure. Indeed, the partial correlation enables the most comprehensive dataset to be compiled on a particular economic subject.[30] Moreover, there is the added benefit that most researchers are familiar with the meaning and interpretation of correlations.

An important drawback of the partial correlation is that, like the simple correlation, its distribution is not normal when its value is close to −1 and +1. For many economics applications this will not be a problem, because no or few partial correlations will be close to these limits.[31] For other cases, the truncation might be a problem, causing an asymmetry on its own. The most common solution to this is to use Fisher's z-transform[32]

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{2.5}$$

This z-transformation also addresses the issue of the standard error of r not being independent of the value of r. At least for the sake of robustness, the meta-analysis can always be conducted using both the partial correlation and Fisher's z-transform. However, in past applications, we have found that these transformations make little practical difference to the central findings of a meta-analysis (e.g. Doucouliagos and Laroche, 2003).

A second drawback of using the partial correlation is that it is not an economic measure of effect. For example, partial correlations cannot be used in environmental economics where the aim of the meta-analysis is benefit transfer. Hence, it might be necessary to supplement the partial correlation with a measure of the economic effect. This will invariably mean that the partial correlation is used for a larger number of estimates than the economic effect. For example, Doucouliagos and Laroche (2009) employed the partial correlation to examine the effects of unions on profits using the results of 45 studies. For 12 of these studies, they also calculated the economic effect, the percentage reduction in profits evaluated at the average degree of unionization.

2.3.4 Elasticities

Elasticity is the most widely used and most commonly known measure of an empirical economic effect. Elasticity measures the percentage change in some important economic phenomenon, say demand or Y, arising from a percentage increase in some stimulus, say price or X. It is a natural effect size to analyze in economics, because there is so much economic discussion about elasticities, and an elasticity is often the crucial magnitude used to gauge the likely effect of a given policy intervention. Many meta-analyses have used elasticities, including Dalhuisen *et al*. (2003) on price and income elasticities of residential water demand, Knell and Stix (2005) on the income elasticity of money, Melo *et al*. (2009) on urban agglomeration economies, and Doucouliagos and Stanley (2009) on minimum-wage employment elasticities. Elasticities are the common effect size used in meta-analysis of research in marketing (e.g. Tellis, 1988; Bijmolt *et al*., 2005; Albers *et al*., 2010).

There are two drawbacks with elasticities. First, the elasticities cannot always be calculated. When the econometric model estimated is in double-log form, the regression coefficients are direct estimates of elasticities. In other functional forms, however, elasticity needs to be imputed from the statistics reported.[33] This is a problem if the authors do not report sample means of independent and dependent variables. While sample means might be approximated/estimated using outside sources, doing so might also introduce a new source of measurement error. In contrast, partial correlations can be directly calculated from routinely reported regression statistics. In some cases, this can make a practical difference. For example, Doucouliagos *et al*. (2012a) explore the links between chief executive pay and firm performance in the UK. The key *theoretical* variable of interest here is the elasticity of executive pay with respect to performance. However, the authors were able to collect only 187 elasticities from 44 studies, compared to 511 partial correlations. Meta-analysis of the elasticities indicated no link between pay and performance. In contrast, the larger dataset of partial correlations suggested a small but statistically significant positive association (a partial correlation of +0.08).[34]

A second hurdle is deriving standard errors. Standard errors are needed in meta-analysis to calculate optimal weights when constructing meta-averages (see Chapter 3), and they are needed to correct a literature for selection bias (see Chapter 4). In a log-log form, this is simple: the regression coefficients are elasticities and their standard errors can be used directly.[35] In other cases, however, the elasticities have to be calculated, usually evaluated at sample means. The standard errors of the regression coefficients in these cases are not the standard errors of the elasticity. The solution to this is to use either the delta method or the Fieller method to estimate the standard error.[36] This means that the number of estimates that can be included will tend to be smaller than if the partial correlation is used.[37]

In the linear form regression (2.1), the elasticity can be evaluated at the mean by calculating $\eta_1 = \alpha_1 \overline{X}/\overline{Y}$, where $\overline{X}$ and $\overline{Y}$ are the average values of the explanatory and dependent variables, respectively.[38] Using the delta method, the variance for this is given by

$$\operatorname{var} \eta_1 = \frac{\overline{X}^2}{\overline{Y}^2} \operatorname{var} \alpha_1 + \alpha_1^2 \frac{\overline{X}^2}{\overline{Y}^2} \operatorname{var} \overline{Y} \qquad (2.6)$$

(see Valentine, 1979, for details). If sample means and variances are not reported, then they might be estimated by using information from outside the study. If this is not possible, then the delta and Fieller methods cannot be applied.

In practice, it is possible to bypass this issue altogether. First, because standard errors are needed to weight the relative importance of each estimate, it is possible to use a different set of weights. For example, instead of using precision, which requires standard errors, it is possible to use sample size or its square root.[39] Second, it might in some cases be possible to approximate the standard error for the elasticity by using the standard error of the regression coefficient, though this might introduce a measurement error. Third, and most important of all, in our view, the imprecision introduced from not using the appropriate standard errors via, say, the delta method, is a second-order concern compared to misspecification and publication selection biases. Most studies that report regression coefficients from which elasticities must be calculated, typically do not report the standard error of the elasticity. If they did, then the issue of deriving the standard errors would no longer be relevant because they would be reported. However, the vast majority of studies do report the statistical significance of the regression coefficient. Hence, the process of selecting which estimates will be reported works through the statistical significance of the *reported regression coefficient*, rather than through the statistical significance of the associated, but *unreported*, *elasticity*. Given that we wish to model the research process and correct any distortions that might arise from it, we see it as more important to use the t-statistics of the regression coefficients, rather than the t-statistics of the elasticity. If publication selection is taking place, it is the t-values of the reported regression coefficients that are being selected. From this t-statistic, t_i, it is very easy to compute a standard error for the elasticity (η), $SE_\eta = \eta/t_i$.

It is worth noting that elasticity concepts vary. For example, while most studies report long-run elasticities, many report short-run elasticities. In this case, researchers may want to use only long-run elasticities or choose to convert all estimates into long-run elasticities.[40] Alternatively, this difference can be modeled in the MRA by including a dummy variable that identifies short-run elasticities (see Chapter 5).[41] The latter approach has the advantage of quantifying the magnitude of the change in responsiveness between the short run and the long run.

There is also the issue of how to treat different concepts of own-price elasticity. Some elasticity estimates include the income effect (Marshallian or uncompensated demand) while others represent the pure price effect (Hicksian or compensated estimates). Some studies will report the conditional price elasticity (demand is conditional upon a subset of the consumer's budget and not the entire budget), while others will report unconditional estimates. Smith and Pattanayak (2002: 285) discuss three approaches to dealing with this issue: (a) pool all estimates and control for differences in the multiple MRA; (b) adjust estimates to "a common economic concept"; and (c) drop estimates that do not use consistent measures. Smith and Pattanayak advocate options (b) and (c). Separate meta-analyses can be conducted for different measures, and they can be jointly estimated using a structural system of MRA equations (Smith and Pattanayak, 2002) or seemingly unrelated meta-regressions (see Chapter 7).

2.3.5 Semi-elasticities

The semi-elasticity measures the percentage change in Y when X changes by one unit. This is a useful measure when the dependent variable is expressed in logs or the equivalent and the explanatory variable is not. Examples include the trade effect of forming a currency union (Rose and Stanley, 2005), the gender wage gap (Stanley and Jarrell, 1998), and the effects of corporate taxation on FDI (Feld and Heckemeyer, 2011). The semi-elasticity has the advantage of not requiring additional information to calculate elasticities. Also, the standard errors for the semi-elasticity are in this case derived directly from the regression output (the standard error for the regression coefficient). It is important to note, however, that only those estimates that use the same scale for the explanatory variable can be combined, otherwise the semi-elasticities are not directly comparable.

2.3.6 t-statistics

Following Stanley and Jarrell (1989), many meta-analysts have used the reported *t*-statistic.[42] Like the partial correlation, this has the advantage of being comparable across estimates and studies, and it can be calculated for all estimates that report a significance level.

However, there are three disadvantages with using the *t*-statistic. First, like the partial correlation, the *t*-statistic is a statistical measure rather than an economic one. Second, it is not as easy to interpret *t*-statistics by themselves. Third, it is also necessary to control for its predictable statistical power. Conventional MRA begins with an effect size that is then converted into a *t*-statistic to correct for heteroskedasticity by weighted least square – see equation (4.2) in Chapter 4. A similar transformation must also be made to the right-hand side (RHS) of the regression by dividing all of the moderator variables by the estimate's standard error (SE) and including $1/SE$ as an additional independent variable (Stanley and Jarrell, 1989; Stanley, 2008) – again see equation (4.2). When specified correctly, this meta-regression model can also identify and correct for publication bias (Chapter 4). The coefficient on 1/*SE* in this weighted least-squares MRA can be interpreted as the estimated effect size. That is, the effect size is measured not by *t*-values but rather by the corresponding partial correlation or elasticity. However, if the *t*-statistic is used as the dependent variable, without a corresponding transformation of the RHS variables, the interpretation of the meta-regression is not what one might expect. Rather, it concerns only the process of publication selection and not the heterogeneity among true empirical effects.[43] This topic is discussed further in Chapters 4 and 5.

2.3.7 Other effect sizes

The most commonly used effect sizes in economics are elasticities, partial correlations and *t*-statistics. There are other measures that might be of interest. For example, Colegrave and Giles (2008) focus on optimal school size, Connor and Bolotova (2006) analyze the size of the cartel overcharge, de Dominicis *et al*. (2008) use the Gini coefficient, and 14 meta-analyses of the value of a statistical life use dollar values (see Chapter 4).

Meta-analysis in environmental economics typically uses dollar values. For example, Simons and Saginor (2006) use the decline in property values from environmental contamination. Fischer and Morgenstern (2006) use the marginal carbon abatement cost. Numerous environmental meta-studies have focused on non-market valuations which are also measured in dollars (Smith and Kaoru, 1990a; Rosenberger and Loomis, 2000a, 2000b; Brander *et al*., 2006).

Meta-analysis of experimental economics has used various measures, such as the cooperation rate in prisoners' dilemma experiments (Sally, 1995), average group efficiency in public good games (Zelmer, 2003), and shares offered in ultimatum game experiments (Oosterbeek *et al*., 2004).

2.4 Coding issues

2.4.1 Research assistants

In our experience, it is not wise to use research assistants to code. The major limitation with research assistants is that they may not be familiar with the literature and, hence, fail to pick up key aspects and technical nuances. A good meta-analysis is not merely an application of statistics. Rather, it is an intelligent and knowledgeable review of an entire empirical research literature, which also happens to contain rigorous statistical analyses to ensure the validity of the insights offered. Hence, meta-analysts must have a deep understanding of the relevant economics literature. The best way to acquire this understanding is to study the literature, inside out. Even after intensive study of the literature, it is still advisable to read the entire papers. Key information may not be reported in tables. Often, essential details are contained in notes to the tables, footnotes, appendices, or "buried" in the middle of the text.

Research assistants are often tempted to just jump to the tables and check off categories on the coding forms.[44] Typically, in economics they do not have their names on the published paper, and hence do not have the same academic incentives for the quality of the final product.

At the minimum, we highly recommend that there are at least two coders to validate the coding. In our own work, we have found it useful to check coding three times. With multiple, independent coders, percentage agreement and other objective measures of reliability, and hence quality, can be calculated. When there is a disagreement about a particular code assigned, it can be checked again to ensure accuracy. Any remaining ambiguity can be arbitrated among the coders.

In our experience, no matter how tightly and rigorously one defines the variables to be coded, rich research literatures will serve up some ambiguity. For example, we coded whether or not the original econometric model included experience in the equation of workers' wages (Stanley and Jarrell, 1998; Jarrell and Stanley, 2004). This seems clear and straightforward, right? Well, how do you code a model that does not have an exact "experience" variable but rather includes worker age, tenure, and age-18? Researchers claim that age-18 is a proxy for experience, "potential experience," assuming that workers' employment is not interrupted. But is this the same as other studies where the data includes an explicit "experience" variable? Does the meta-analyst accept the claim made in the primary literature that age-18 is an acceptable proxy for experience? Of course, we could have two "experience" variables: one that codes for whether there is a distinct variable for experience, and a second for potential experience. However, this is a slippery slope. Having such fine-grained codes for all potentially relevant research dimensions is not feasible. Academic economics rewards innovations; thus, there are usually a greater number of combinations of model, data, and technique variations that one can find in a research literature than estimates. Often there are not sufficient degrees of freedom to code everything; thus, some choice about what really matters will need to be made *a priori* by the meta-analyst.

2.4.2 Locating the data

In economics, most of the data on effect sizes (regression coefficients, standard errors, *t*-statistics or *p*-values) will be reported in tables in the main text of a study or in an appendix. Increasingly, some of the data is available via web addresses. Much information, however, will also be buried in the text of the paper, including important information on the measurement of the variables and the estimation technique.

At times, the information may be presented in figures. For example, some studies report the results from vector autoregression models in the form of impulse response functions. While there is some degree of measurement error involved, it is possible to extract precious data from the graphs themselves, converting plots into effect sizes. Two meta-analyses of the impact of monetary policy have done exactly this (Ridhwan *et al*., 2010; Rusnák *et al*., 2011).

2.4.3 Missing information

Reporting standards vary between journals and over time. As already noted, we require, at a minimum, data on an effect size and its standard error. Effect sizes are rarely reported in the manner we desire them by all studies; much of the time partial correlations or elasticities have to be calculated from the reported statistics. Fortunately, *t*-statistics of the regression coefficient are usually reported. As long as the estimated regression coefficient and either its standard error or *t*-value are reported, the other statistics are easily calculated from $t = a_1/SE_{a1}$, where a_1 is the estimated coefficient. However, we have a serious problem if neither the standard error nor the *t*-statistic is reported.

Where the exact *p*-values and degrees of freedom are reported, it is possible to work backwards to derive the *t*-statistics, exactly.[45] In other cases, only the level of statistical significance is reported, for example with *, **, and *** denoting statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. In this case, the meta-analyst needs to decide whether to include these estimates/studies at all. If the estimates are to be included, their standard errors will need to be imputed from the *t*-value. However, to do so further requires that the meta-analyst takes one of four approaches, and all four introduce some measurement error into the meta-data.

The simplest is to assume that an estimate that is statistically significant at the 1 percent level has a *p*-value of 0.01. The *t*-statistic can then be established from this assumption. Likewise, if a variable is statistically significant at the 5 percent level, one might assume that the *p*-value is exactly 0.05. The second approach is to follow Greenberg *et al*. (2003) and assume that the actual *p*-value lies at the midpoint of statistical significance range. Thus, an estimate that is significant at the 10 percent level is assumed to have a *p*-value of 0.075, an estimate that is significant at the 5 percent level is assumed to have a *p*-value of 0.03, and an estimate that is significant at the 1 percent level is assumed to have a *p*-value of 0.005. The third option is to use the distribution of estimates from those studies that have reported exact *p*-values (or from those that reported sufficient information to calculate exact *p*-values) and assume that the studies that report incomplete information follow the same distribution. Fourth, we can omit these estimates altogether, but this reduces the dataset.

Lastly, there might be an issue of how to handle estimates where the author reports a coefficient and states that the estimate is not statistically significant, without reporting any further statistics. Usually, such estimates are omitted from the meta-analysis, and this is probably the best thing to do because any imputation is likely to introduce a bias into the meta-analysis, perhaps a large one. If they are to be retained, one approach is to assume that the *p*-value was 0.10, on the basis that authors will try to get *p*-values as close to the 10 percent level of significance as possible. Alternatively, a *p*-value of 0.5 can be assumed, but, of course, this is just a wild guess. Greenberg *et al*, (2003) use 0.3, as this is the midpoint between 0.10 and 0.5.

2.4.4 Multiple estimates: all, best, independent and average datasets

There is always an issue about how to handle multiple estimates reported in a given research study. Multiple estimates are much more common in economics where editors and reviewers demand that applied econometric studies report multiple models, methods and estimates to ensure the robustness of the authors' main findings. In Chapters 4 and 5 we discuss statistical approaches to accommodate the potential statistical dependence that might be lurking among multiple estimates in the same study. However, there are also alternative approaches to collecting the data that can remove the issue of within-study dependence. It is to these alternative ways of defining the meta-dataset that we now turn.

The *best-set* of estimates consists of one estimate from each study, using the key regression from each paper. Ideally, it is the one that is explicitly preferred by the authors themselves. Unfortunately, it is not always clear what the authors' preferred estimate is, so it becomes necessary for the meta-analyst to make some judgment. In the absence of publication bias, the best-set might be preferred. However, given the strong statistical evidence of widespread publication selection (Doucouliagos and Stanley, 2012), it is possible that the authors' preferred estimate reflects greater selection than other reported estimates. Hence, the "best-set" might not be the *best* dataset to use in a meta-analysis.

The *average-set* is constructed by taking an average of all effect sizes reported by each study. Ideally, this should be a weighted average using optimal weights.[46] Stanley (2001) recommends this approach to intra-study dependence, and Krueger (2003) finds that when followed it makes a large difference to the overall assessment of the effect of class size on student achievement. The disadvantage of the average-set is, however, that it fails to take advantage of potentially relevant information. The within-study variation can be very informative.

The *all-set* consists of all relevant estimates reported in each of the studies. This often greatly increases the number of observations available for meta-analysis, though it does result in added potential interdependence between data points, which needs to be accommodated by appropriate statistical methods. The advantage of using the all-set is that it offers more estimates to explain the large variation (heterogeneity) typically found between studies and between estimates. Furthermore, it does not contribute to selection bias, potentially inadvertently introduced by the meta-analyst herself.[47]

The *independent-set* of estimates consists of only those estimates that are deemed to be *conceptually independent*. Following Hunter and Schmidt (2004), a study can be regarded as conceptually independent, in this context, if it uses the same dataset as a previous study but involves different authors, or if the same authors use different datasets. For example, some studies might report the effects of an independent variable on the dependent variable for different groups of countries, such as for the OECD, for African developing countries, and for Asian developing countries. These could all be treated as independent as they all use different samples.[48]

Many meta-analysts prefer to use the average-set, following Stanley (2001). Some use two or more of the datasets. Multiple meta-datasets allow results to be compared and their robustness explored. Conventional practice has evolved to use the all-set as the standard dataset and to model potential within-study dependence with multi-level, unbalanced panel and cluster-robust MRAs. For the sake of robustness, it is also advisable to use one of these other datasets. The average-set is a natural extension of the *all-set*, because it must be computed from the all-set.

2.4.5 Systematic versus partial reviews

The greater majority of meta-analyses have been systematic. That is, they have sought to identify the population of estimates for a particular literature, such as the effect of *X* (advertising) on *Y* (alcohol consumption), and proceeded to pull together all of this data. The studies reporting on *X*'s effect will typically also report estimates of other effects, say *Z* (price of alcohol) on *Y*. These estimates can also be coded and a meta-analysis can then be conducted upon the effect of *Z* on *Y*. While the meta-analysis of literature on *X*'s effect may be regarded as systematic, the meta-analysis of the literature on *Z*'s effect is only partial. Only that part of the literature that explores the effects of both *X* and *Z* is surveyed and analyzed. The effect sizes from the partial reviews of, say, *Z* on *Y* can then be compared to the effect sizes from the systematic reviews of *X* on *Y*. This can help to compare the relative importance of the main effect size that is under investigation. Examples of this approach can be found in Doucouliagos and Ulubasoglu (2006) and Doucouliagos and Laroche (2009).

2.5 The quality conundrum: should estimates be combined?

> If certain studies are to be ruled out as being of low quality, meta-analysis impels the researcher to enunciate and code the specific characteristics that identif[y] the inferior nature of these studies. Studies should be omitted only by objective criteria that are applied evenly across the entire literature.
>
> (Stanley, 2001:147)

Meta-analysts typically try to be as comprehensive and inclusive as possible so as not to distort their findings. Studies can easily be identified, their characteristics and estimates can be coded, but does it make sense to combine estimates from different studies? Studies differ in many respects. An oft-made criticism of meta-analysis, especially by referees new to meta-analysis, is that it is inappropriate to combine the results from different studies when studies differ in quality. Differences in the quality of studies might result in biased estimates and invalid inferences, or so the old chestnut goes. In our view, this tired argument, repeated by referees, is a "red herring." "Quality" is often in the eye of the beholder and can be a thin chemise used to cover naked bias for one's own theory or for the conventional theory. Vague notions of quality, ascribed to some selected methodological distinction, have often been used to select only those results that fit into one's preconceived views (Stanley, 2001).

Study quality differences can be perceived or real. Perceived differences in study quality often take the form of a bias in favor of ranked or leading journals. For example, many would subjectively expect that a study published in the *American Economic Review* will be of higher quality than a study published in, say, in a small regional journal. The issue for the meta-analyst, however, is whether there are more objective measures of quality, such as the precision of the estimates in question. Doucouliagos and Stanley (2008) compare the precision of over 12,000 estimates against various measures of journal quality. No systematic differences were detected – higher-quality journals did not report estimates that possess greater precision. One explanation for this is that leading journals focus more on the quality of the narrative and the introduction of new data, estimators and methods. The precision of the estimate is often not considered to be the key contribution nor a focus of papers published in these journals.[49] Hence, it would be inappropriate to dismiss *a priori* studies because they are published in a less highly ranked journal.

As part of the coding of the studies, we recommend that information is collected on study quality. Several measures of quality are available. First, we can use each estimate's precision as the indicator of quality. This is the most statistically valid approach, as it is derived directly from the study's estimate and does not rely on any additional judgment. Precision is calculated as the inverse of the estimate's standard error. More precise estimates can then be assigned a higher weight. Typically, this means that smaller studies are given less weight, as these tend to report estimates with less precision (see Chapter 3 on summarizing meta-analysis data). Second, we can use the impact factor of the journal in which the study was published. The impact factors are reported in the *Social Science Citation Index* (SSCI). Journals with larger impact factors might, arguably, be considered to be of higher quality and, hence, assigned a larger weight. A disadvantage to this approach is that impact factors are not available for all journals.[50] Third, the number of citations each study has received, as reported in the SSCI (or Scopus), may be considered a "revealed quality preference."

The impact factor rankings are based on a measure of the quality of the journals in which the studies were published, rather than some measure of the quality of the studies themselves. The citations data are based on the studies, rather than the journals. The explicit assumption made is that studies that receive more citations should be assigned more weight, as they have been more influential. Journal impact factors and the number of citations can be considered to be measures of what the profession deems to be important. We prefer the more traditional approach in meta-analysis, which is to use the estimate's precision (Hunter and Schmidt, 2004; Stanley and Doucouliagos, 2010; Stanley *et al*., 2010). We take the view that precision is the better weight, as it is based on objective statistical information alone. A key advantage of precision is that it will be available for all estimates included in our meta-data.[51]

Some meta-analysts only use estimates from the leading journals. In our experience, this rarely makes a difference to the overall meta-analysis results. However, it might be important to referees not already part of the meta-community. Most studies that have explored this issue find that there are no discernible differences in the estimates. Examples include Klomp and de Haan (2010) and Disdier and Head (2008). In contrast, Gallet and List (2003) find differences in both price and income elasticities between leading and non-leading journals.[52]

Where referees insist that only estimates from "leading journals" should be used, or in anticipation of such a criticism, we recommend three options. First, a simple regression of precision (the inverse of the standard error) against journal quality can be run:

$$\text{Precision}_i = b_0 + b_1\text{JournalQuality}_i + u_i \tag{2.7}$$

If b_1 is not positive and statistically significant, then there is no evidence that "leading journals" report more precise estimates. Hence, there is no need to focus only on the estimates from these leading journals, and there is no objective reason to discard the information contained in other journals.

Second, the meta-analysis can be conducted for all estimates and then repeated with only the estimates from "leading journals". Third, measures of study quality can be included in the meta-regression models as another potential explanatory variable.

In the majority of cases, perceived notions of journal quality and journal rankings are unlikely to be an important conditioning factor. There is, however, one other valid and important dimension of study quality: methodological rigor. Studies do differ widely with regard to specification, data, estimators, etc. The great thing about meta-analysis is that, by combining and coding these differences, the analyst is able to quantify objectively and rigorously the effect of these observable dimensions of quality on the reported estimates. Then the analyst is able to infer "best practice" for the research literature in question.

Of course, meta-analysis can be conducted on only those studies that are deemed to be "best practice." This could be part of the data collection strategy. However, why omit studies on questionable methodological grounds when such observable methodological differences can be modeled explicitly in multiple MRA? In this way, even the poorest studies, however defined, can help to identify reported variations in effects due to differences in methods, data, and techniques. From a pure statistical point of view, we should include all relevant estimates and code for all observable differences in quality and methods. Then we can let the research record itself reveal the importance and the effects of these research dimensions on the reported findings. Our own preference is to cast the widest net and test whether differences in quality or rigor make a difference. It is often the case that what is conventionally perceived to be an important dimension of "best practice" makes little practical difference to reported research, once the other dimensions of the research process are fully incorporated.

2.5.2 Data dependence

When more than one estimate per study or per author is employed, the meta-analyst might encounter data dependence. That is, estimates reported within a single study might not be statistically independent of each other. Such data dependence can take three forms:[53]

- *Study dependence*. When studies report more than one estimate, estimates are not strictly independent of each other.
- *Author dependence*. If authors publish more than one study, estimates between these studies may not be independent of each other.
- *Spatial dependence*. When researchers receive direct feedback from each other or are influenced by prior findings, this might cause data dependence.

In other areas of research (e.g. medical research and psychology) using experimental trials, one estimate is likely to be independent of the next. In economics, meta-analysis is often criticized on the grounds that the data are not independent. Yet in some cases, they are. For example, when looking at the effects of unions on productivity, studies often sample entirely different establishments. Similarly, experimental economics uses entirely different samples of (typically) student subjects. However, in many other areas of economics research, data dependence/independence is likely to be more complex. Data dependence is particularly a problem to the meta-analysis of macroeconomic data. Most applied econometric studies draw from the same data sources (e.g. World Bank Development Indicators, Penn World Tables, US Bureau of Economic Analysis, etc.), which implies that the statistical results from the same data should be somewhat related.

So, where studies use the same data, is meta-analysis meaningful? If the studies cannot logically be combined in a meta-analysis, then they cannot be combined in a traditional qualitative literature review either. This would then mean that it would not be possible to draw inferences from macroeconomic studies. However, unlike conventional narrative reviews, the meta-analyst has two objective strategies at her disposal.

First, it is routine to code for several of the dimensions of the original research, including those potentially related to this dependence. For example, while all studies might be drawn from the World Bank Development Indicators, they may not include the same set of countries and time periods, and some may make independent misspecification errors through the researchers' idiosyncratic choices of exact model specifications and methods. MRA routinely controls for the country composition of the samples, the time periods used and other potentially dependent dimensions of the research results.

Second, this problem of dependence is likely no worse in meta-analysis than in macroeconomic research in general. Stanley and Jarrell (1989) argue that the issue of data dependence is likely to be less problematic for MRA than for the typical econometric applications where autocorrelation, path dependency, and non-stationarity are ubiquitous and strong. Such potential dependence is routinely handled in applied econometrics by more sophisticated regression models or techniques. The same set of tools, methods, and models available in conventional econometric studies are also available to the meta-analyst. Thus, if dependence is thought to be a further concern, we can use more sophisticated regression techniques to handle it. It has become standard practice among meta-analysts to employ multilevel (or unbalanced panel) MRA models to account for potential dependence explicitly and cluster-robust standard errors to correct for its potential effects. These topics are discussed in greater detail in Chapters 4, 5, and 6.

2.6 Summary

The coding of research is by far the hardest and most time-consuming step in a meta-analysis. However, it is important because it provides the raw material for meta-analysis. When coding, the meta-analyst needs to be as inclusive, comprehensive, yet insightful, objective, and as rigorous as is practically feasible. Undertaking a meta-analysis is no shortcut to the “truth” or to an easy publication. But it is likely to be the only path to a genuine understanding of contemporary research in economics, business, social science and medicine.

A good meta-analysis is both an insightful, but short, narrative review and a comprehensive and rigorous econometric analysis of the full research record. When in doubt, we encourage meta-analysts to err on the side of being:

- inclusive in the research results collected;
- comprehensive in identifying and coding differences in research methods, data, and models employed that might potentially explain the large variation observed among reported research results;
- objective in defining clear criteria for study inclusion/exclusion and for coding variables so that the resulting meta-analysis is independently replicable, which is the hallmark of science (Popper, 1959);
- insightful and creative in identifying factors which might drive the reported research;
- transparent and explicit about how studies were selected and coded.

Large amounts of scarce intellectual and financial resources are invested in producing economics research. It would be a great waste to fail to glean the few nuggets of knowledge contained in our mountains of research results. Econometrics provides the necessary tools needed to refine this raw research ore, but it is up to economists and business researchers to take the time to employ these tools on research itself.

3 Summarizing meta-analysis data

In the previous chapter we discussed how to search and code research for a meta-analysis. This chapter focuses on the description of these data, presenting alternative ways of summarizing research findings. It is important to note at the outset that describing a literature is not the central aim of meta-analysis in economics. Meta-analysis offers so much more than this. The main contribution of meta-analysis is to make inferences about the state of economic and business knowledge and to correct a literature for misspecification and selection biases that typically plague empirical studies (see Chapters 4 and 5). This more analytic and comprehensive meta-analysis enables meta-analysts to test rival theories and to provide accurate and corrected estimates of policy-relevant parameters.

While not the central focus of meta-analysis, it is nevertheless extremely useful to commence a meta-study with a close look at the data. We will illustrate all aspects of meta-analysis, including descriptive summaries, using data from four published meta-analyses: the effects of unions on productivity (Doucouliagos and Laroche, 2003), residential water price elasticities (Dalhuisen *et al*., 2003), the value of a statistical life (Bellavance *et al*., 2009) and minimum wage elasticities (Doucouliagos and Stanley, 2009). Table 3.1 summarizes some of the key features of the data collection for each of these studies, and the nature of the meta-analysis. The interested reader should refer to these studies for further details.

3.1 Illustrating data

Once the meta-data has been coded and the effect sizes calculated, the meta-data can then be analyzed. While most meta-analyses are presented without descriptive statistics, in our experience it has been very useful to commence with graphs.

Several types of graphs have been used in the meta-analysis of empirical economics. Some are graphs that are widely used in statistics. Examples include simple frequency distributions of either *t*-statistics or effect sizes (e.g. Doucouliagos and Laroche, 2003; De Mooij and Ederveen, 2003; Bijmolt *et al*., 2005; and Holmgren, 2007), box and whisker plots (e.g. Smith and Huang, 1995; Brander *et al*., 2006), and stem and leaf plots (e.g. Verlegh and Steenkamp, 1999).[1]

A second group of graphs is more specific to meta-analysis, such as funnel graphs, forest plots, Galbraith diagrams, and L'Abbé plots (see Sutton *et al*., 2000).

Table 3.1 Four illustrative meta-analyses

Field	*Search engines used*	*Type of data included*	*Effect size*	*Number of studies (estimates)*	*Countries studied*	*Aim of study*	*Explores publication bias?*
Unions and productivity	Numerous	Published in English and French	Partial correlation	73 (73)	Various	Testing theories	Simple test
Water demand	Numerous	Published and unpublished	Elasticity	50 (110)	Various	Parameter estimate	No
Value of a statistical life	Numerous	Published only	Dollar value	37 (39)	Various	Parameter estimate	No
Minimum wage	Numerous	Published and unpublished	Elasticity	64 (1,474)	USA	Testing theories	Extensively

We have found the funnel graph to be the most useful of these (see Stanley and Doucouliagos, 2010). The funnel graph does a nice job of displaying publication selection bias, which is discussed in detail in the next chapter. Funnel plots and frequency distributions appear to be the most common graphs used in economics meta-analysis. Forest plots can be used to illustrate the distribution of estimates by plotting each estimate and its associated confidence interval. They show both the pooled mean as well as the variation (Lewis and Clarke, 2001). Examples in economics include Capelle-Blancard and Couderc (2007) and Havránek (2010).

3.1.1 Funnel graphs

A funnel graph is a scatter diagram of all empirical estimates of a given phenomenon against these estimates' precisions (i.e. the inverse of the estimates' standard errors, 1/*SE*).[2] A clear example of the expected funnel shape can be seen among econometric studies of the productivity effects attributed to unionization (Figure 3.1). The main use to which funnel plots have been applied is to illustrate publication bias in a literature. However, funnel plots are also a very useful way to identify coding errors, outliers and potential leverage points in a literature. They may also be used to identify heterogeneity. Further, they also show, rather vividly, the wide variation in reported empirical results. In the case of Figure 3.1, we can see clearly that there are many negative and many positive results reported and that there is also a large cluster of observations around a zero effect. Using funnel graphs to identify heterogeneity and publication selection bias is discussed in detail in the next two chapters. Here we focus on the role that funnel graphs can play in double-checking the accuracy of our meta-data.

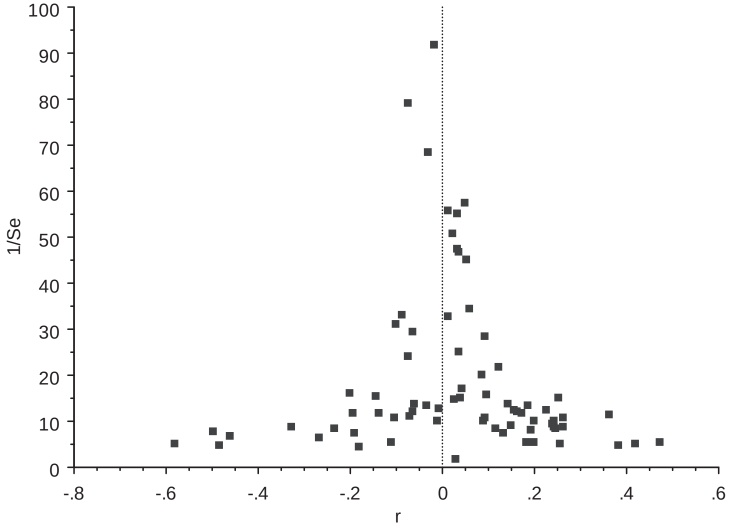


Figure 3.1 Funnel plot of union-productivity partial correlations

Source: Doucouliagos *et al.* (2005).

3.1.2 Detecting coding errors

In a handful of cases, we have correctly identified coding errors by merely looking at the funnel graph. As the name suggests, the funnel graph should more or less resemble an inverted funnel (see Figure 3.1).[3] Publication selection may distort such a graph by removing many of the points on one side or the other. The implausibly shaped funnels that we have identified come in two forms. In a few cases, we have seen very high points, near the center, but very much higher than any other estimate in the literature. Every time we have observed this, double-checking the original papers uncovered that either the unusually large precision in question was in fact an error or the estimate was not comparable to the other studies. In one case, the mistake was as simple as adding an extra "0" to an already very small standard error (of the order of 0.001).

The second way that a funnel graph discovers errors happens when there are one or two points with relatively large precision, but not necessarily the largest, and the associated estimates (measured on the horizontal axis) are much different than the center of the funnel, as defined by all the other estimates in the literature. In other words, one or two estimates are far to the right or to the left of the rest of the funnel. Although this may also be a sign of genuine heterogeneity and no mistake, it is nonetheless worth rechecking the coding. Obviously, if it is an error, the error should be corrected. All meta-data points should be checked, rechecked and verified. It is impossible to be too meticulous in validating the accuracy of one's codes. If the unusual point of the funnel graph is correct, it is still useful to reread the paper and the other codes to see if there is something genuinely unique about this estimate. If the estimate is coded accurately, then there must be some research factor or dimension that explains this precise but very different value. If this research point is correct, and no research factor explains this difference, then our MRA will have a large amount of unexplained heterogeneity, which can invalidate its summary findings.

3.1.3 Detecting outliers and leverage points

The funnel plot can also be an excellent way to detect outliers and leverage points. Simply stated, outliers are extreme and implausible values of the dependent variable, and leverage points are extreme values of the independent variable that can exert great influence, even when correct, on the regression (or MRA) relation. The definitions of independent and dependent variables come from the conventional MRA model (see equations (4.1) and (5.5) in the following chapters), where effect size is the dependent and its standard error is the independent variable.[4] Deciding what to do with outliers can be rather nuanced and is likely to be different when applied to funnel graphs than to the raw economic data used in econometric research. We can distinguish between two potential types of "outliers." Some outliers might involve effect sizes with low precision but very large values of the estimates, either positive or negative. For example, union-productivity correlations larger in magnitude than 0.4 or so might be seen as "outliers" (Figure 3.1). These relatively large but imprecise effect sizes, whether technically classified as outliers or not, can be retained in our database with little or no harm to (or effect on) our results. Standard meta-analysis is conducted using some function of precision as weights (see the next two chapters); thus, these very imprecise outliers will exert very little influence on any of the meta-analysis results.

On the other hand, leverage points defined by having high precision can have a correspondingly large effect on the results of meta-analysis. As discussed above, unusually large precisions may be a sign of a coding error. When not in error, however, they must be retained because they are genuinely informative about the research literature in question. Such large precisions are not really outliers, but rather leverage, or influential, points. Unless a valid and independent reason for the removal of these leverage points can be found, such as that they come from a distinct population or use a unique measure of effect size, they should be retained in the meta-data. Robust meta-regression techniques are always a wise choice in MRA and will minimize the undue influence of any one or few values in the literature.

3.1.4 Chronological ordering of the data

The graphical representation of meta-analysis data using a chronological ordering may offer additional insight, as it can capture the evolution of the literature. A key benefit of such ordering of the data is that it can trace the evolution of the effect sizes, highlighting trends and possibly structural breaks in a literature. As an example, consider Figure 3.2, reproduced from Doucouliagos and Paldam (2008). The data here are partial correlations of the effect of development aid on economic growth. The horizontal axis shows these reported effects in chronological order. A simple linear trend line can be fitted to this area of research and represents a statistically significant *decline* in the reported size of the effect of aid on growth. This downward trend potentially has an important economic interpretation. Doucouliagos and Paldam (2008, 2009) argue that after 40 years of development aid assistance, aid agencies should be getting better at choosing successful aid projects. That is, partial correlations should be rising over time, rather than falling. These authors point out that this declining trend may be explained by publication bias. Researchers are reluctant to show that aid does n'ot work, but, as more data accumulated, the reported estimates tend to converge to their actual underlying empirical effect, which by several measures seems to be zero.[5]

Such chronological orderings might not make sense in all fields. However, this chronological graph may also help to identify a genuine trend in the underlying economic phenomenon studied.[6] For example, several meta-analyses have found a clear trend among male–female wage differentials (Stanley and Jarrell, 1998; Jarrell and Stanley, 2004; Weichselbaumer and Winter-Ebmer, 2005). Changing social attitudes about gender roles and discrimination laws predict that there would

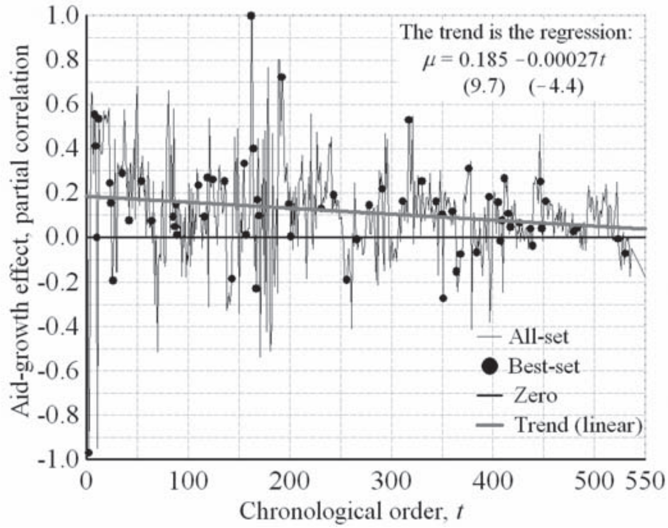


Figure 3.2 Chronological ordering of data

Source: Doucouliagos and Paldam (2008).

be a decline in the actual amount of gender wage discrimination. Doucouliagos and Stanley (2009) present a similar diagram for the minimum wage literature, illustrating a significant linear trend showing a lessening of minimum wage's impact over time (0.04 less negative every decade). Although this chronological graph can help to identify trends and path dependencies among research findings, such effects can also be captured by multiple MRA. It has become routine to include trend variables in multiple MRAs (see Chapter 5).

3.2 Summary measures

Many meta-analyses begin with a description of the distribution of the empirical results. This sets the baseline of the typical reported effect size in the research literature in question and provides a context in which to understand more sophisticated multiple MRA. Conventional summary statistics such as the weighted mean (fixed effects and random effects), median, standard deviation and frequency distribution should always be reported. Before looking at these, we first consider the usefulness of vote counting and integrating p-values.

3.2.1 Vote counting

It is tempting to assess a literature by simply *counting* estimates and then, for example, comparing the number of estimates that are positive to those that are negative, or to compare those that are statistically significant to those that are not. This vote-counting exercise is presented in Table 3.2 for the four meta-analyses.

Table 3.2 Vote counting

Field	*Negative and statistically significant*	*Negative and statistically not significant*	*Positive and statistically not significant*	*Positive and statistically significant*
Unions and productivity	26%	14%	28%	32%
Water demand	83%	13%	3%	2%
Value of a statistical life	0%	0%	10%	90%
Minimum wage	42%	32%	18%	7%

One use of vote counts is that they present a rough summary of the distribution of findings and the extent of apparent disagreement within a field. For our four datasets, it appears that unions and productivity has the most disagreement. For water price elasticities, it appears that there is agreement regarding a negative effect. Readers and reviewers routinely carry out such vote counts by selecting studies with conflicting estimates to illustrate that there is disagreement within a given literature.

While fairly straightforward, vote counting has several disadvantages, all of which stem from the deliberate loss of information; vote counting essentially means taking a distribution of estimates and collapsing this into two to four categories. The resulting loss of information is unnecessary and often misleading.

Vote counting can create problems where there were none. First, as Hunter and Schmidt (2004) note, vote counting can be misleading because the vote counts ignore the effects of sampling error. Sampling error creates variation in results that makes them look like they disagree. When 95 percent confidence intervals are constructed around the estimates, an entirely different conclusion might emerge.

A second problem with vote counting is that it does not provide an estimate of an economic magnitude, such as an elasticity, that can be used for policy making. Statistical significance is an important first step, but insufficient by itself. Providing reliable estimates of economic parameters is critical for both policy and economic understanding.

A third problem with vote counting is that by taking away the focus from elasticities, it also takes away the focus from publication bias. We show in the next chapter that the union-productivity literature is relatively free of selection bias, whereas the literature on water price elasticities, the value of a statistical life, and minimum-wage effects are highly contaminated by it. By focusing on vote counts, the meta-analyst will entirely miss a key structural weakness in the data and the need to correct the data before valid inferences and sensible policy recommendations can be drawn.

A fourth problem is that we frequently require a clear understanding of the source of variation between studies – not just whether estimates differ in terms of the level of statistical significance. There are several cases of meta-analysis that begin with vote counting and then proceed to use the vote counts as the dependent variable in the MRA. As noted in Chapter 2, such meta-probit (or meta-logit) models can be highly problematic.

Lastly, vote counts have been shown to possess a statistically perverse property. Due to the low power of many statistical tests, Hedges and Olkin (1985) show that the probability that a vote count comes to the wrong conclusion actually increases as research accumulates. It is an understatement to suggest that this is less than the statistical ideal. Our recommendation is that vote counts be used very sparingly or not at all. Although vote counts could be used as an alternative way to illustrate the distribution of the meta-analysis data, they lose much information and thereby are less useful than a simple funnel graph. Compare Table 3.2 to the relevant funnel graphs in Chapter 4.

3.2.2 p-values

Arguably, meta-analysis began in the early twentieth century when R.A. Fisher and Karl Pearson independently developed procedures to summarize the overall effect of multiple independent tests (Fisher, 1932; Pearson, 1904). Some meta-analysts still use Fisher's method of combining *p*-values to see whether a research literature, when seen as a whole, shows a statistically significant effect. Under the null hypothesis that there are no genuine effects, *p*-values (P_i) will be uniformly distributed:

$$f = -2\sum_{i=1}^{L} \ln P_i \overset{d}{=} \chi^2\,(2L) \tag{3.1}$$

for a literature containing L studies. Fisher's test is very generous in ascribing statistical significance; hence, its popularity. In order for the Fisher test for an overall effect to be valid, the research findings cannot have heterogeneity or biases, conditions that are rarely, if ever, met in economics and business research.

The problem is that the underlying assumption for Fisher's test is that the values being estimated are all exactly zero. When different studies use different countries, time periods, estimation techniques, or independent variables, this assumption is very likely to be invalid. Some of these differences will produce genuine effects (heterogeneity), even when the overall effect is actually zero. Also, many of these variations in methods, models, and variables will produce non-zero biases. Unfortunately, it takes only one bias for the null hypothesis of Fisher's test to be literally false and for the calculated test statistic to become statistically significant when large enough. Thus, rejecting the null hypothesis of the Fisher test does not actually mean that there is a genuine empirical effect, as it is usually interpreted, but rather that there are either biases or heterogeneity. Unfortunately, both of these are known to be common in economics and business research.

Worse still, if some of the studies in a given area of research select statistically significant estimates to report, then Fisher's test is virtually guaranteed to give an indication of genuine empirical effect even where there is none. In the next chapters, we discuss the commonplace nature of misspecification and selection biases in economics research and how meta-regression analysis can identify, accommodate, and correct these biases. In sum, Fisher's test tests the wrong hypothesis (that all estimates come from a population with a zero mean) and is quite likely to be misinterpreted. Consequently, we recommend that *p*-values are not combined.

3.2.3 Descriptive statistics

A simple (unweighted) average is often reported to summarize the findings from a literature. However, the weighted average effect size, say an elasticity, η_w,

$$\eta_w = \frac{\sum w_i \eta_i}{\sum w_i} \tag{3.2}$$

is statistically the preferred choice, where the w_i are the weights used, and η_i is the measure of the estimated elasticity. The optimal weights have been shown to be the inverse of the estimates' variances (Hedges and Olkin, 1985; Cooper and Hedges, 1994). However, optimal weights might not always be practical. Hence, other weights might at times be used. For example, the sample size can be used where standard errors and, hence, variances are not available;[7] journal impact factors can be used as a measure of the quality of the journal in which the study was published; citations can be used as a measure of the importance the profession has placed on the study; and for survey-based studies, it might be possible to use the survey response rate. Some authors use weights that take into account the number of effect sizes included from each study, in order to ensure that no one study exerts an undue influence on the results. While the inverse of the variance is optimal from a statistical point of view, these other weights can be used as a sensitivity analysis or as a matter of practical necessity.[8] Nevertheless, preference should be given to the use of optimal weights.

The estimated standard error of η_w can be calculated as the square root of the reciprocal of the sum of the optimal weights. This can be used to construct confidence intervals and to test the statistical significance of the weighted mean.[9]

Averages can also be calculated for subsets of the data. For example, for sensitivity purposes, we might want to report the weighted mean for certain regions (or firms), certain time periods, for specific measures of effect, or for those estimates published in what are deemed to be leading journals.

Fixed versus random effect estimates

There is much discussion in the meta-analysis literature about fixed- and random-effects estimators (FEEs and REEs, respectively).[10] FEEs weight each reported estimate by the inverse of the square of its standard error (or equivalently its precision squared). FEEs assume that all of the reported estimates are drawn from the same population with a common mean. When estimates are drawn from several populations (i.e. when there is heterogeneity), the REE becomes, technically, the appropriate estimator. REEs weight each estimate by the inverse of a more complex variance that contains two components, $SE_i^2 + S_h^2$, where S_h^2 is an estimate of the between-study or heterogeneity variance.

The unweighted and weighted averages and associated 95 percent confidence intervals for our four datasets are reported in Table 3.3. Note that the unweighted average is greater than the weighted averages, with the exception of unions and productivity. A simple unweighted average will in most cases give a misleading measure of the effect size. Note also that the fixed-effects estimate often differs a lot from the random-effects estimate, in some cases significantly so. The weighted averages and the 95 percent confidence intervals suggest that: unions have no effect on productivity; water is price inelastic; the value of a statistical life is positive; and minimum wages have an adverse effect on employment.

Unfortunately, these conclusions are premature. The weighted effect will be an unbiased estimate of the population effect, as long as the studies included in the calculation are all the available estimates, *or* a random sample from the population of all estimates (see Hunter and Schmidt, 2004). Where publication selection bias is present, all averages, weighted or not, are distorted (Stanley, 2008). Even though the REE is the proper weighted average to use when there is excess heterogeneity, likely publication bias reverses this conventional judgment. Simulations show that the FEE is less biased in the presence of publication selection (Stanley, 2008; see also Chapter 4 below). We show in Chapter 4 that with the exception of the union-productivity effects, these literatures suffer from significant publication selection bias, the effect of which is to distort all averages. For example, publication bias greatly inflates the value of a statistical life, and it also gives the impression that minimum wages have an adverse effect on employment when they have no employment effect (Chapter 4). Meta-analysts should refrain from drawing any inference from these averages, weighted or unweighted, unless publication selection is formally tested and found to be absent.

Simple linear regression

An alternative approach is to run the following simple linear regression:

$$effect_i = \beta_0 + u_i \tag{3.3}$$

Table 3.3 Unweighted and weighted averages

Field	*Unweighted average*	*FEE*	*95% CI (FEE)*	*REE*	*95% CI (REE)*
Unions and productivity	+0.021	−0.0003	−0.009 to +0.008	+0.023	−0.0009 to +0.0463
Water demand	−0.378	−0.116	−0.12 to −0.11	−0.29	−0.32 to −0.27
Value of a statistical life (US$ million)	9.5	1.8	1.6 to 1.9	5.7	4.6 to 6.8
Minimum wage	−0.191	−0.037	−0.03 to −0.04	−0.105	−0.11 to −0.10

This has much appeal, as it falls naturally within the regression framework adopted by most empirical economics, and it can be a springboard for much more complex multiple MRAs (see Chapters 4 and 5). Equation (3.3) is a fixed-effects model that assumes that the effect sizes vary randomly around a single value, β_0. If all estimates are to be treated equally and given an equal weight, then equation (3.3) may be estimated using ordinary least squares. If estimates are to be treated unequally, then weighted least squares must be used. For example, using the inverse variance as weights gives the FEE. In either case, $\hat{\beta}_0$ is the estimate of the effect size and the associated *t*-statistic provides a test for whether this is statistically significant. Random-effect weighted averages can be calculated from a regression model that adds an independent random term to equation (3.3). The next three chapters discuss more sophisticated versions of these models in greater detail.

3.3 Statistical significance versus economic significance

The calculated weighted average from equation (3.2) provides an estimate of the underlying parameter of interest. These estimates can be thought of as *meta-averages* but should be interpreted with caution because of likely biases and heterogeneity. Furthermore, McCloskey (1985, 1995) highlights the importance of economic significance, as opposed to statistical significance. That is, instead of just focusing on the statistical significance of any meta-average, it is important to note also the economic meaning, if any, of the size of the estimated effect. Some meta-averages might be statistically significant, but so small as to be of little economic meaning. This is precisely what we find among the nearly 1,500 estimated employment effects of minimum wages (Doucouliagos and Stanley, 2009). The fixed-effects estimate of the minimum-wage elasticity of employment is −0.037.[11] Although this is statistically highly significant ($t = -16.6$; $p < .001$), it is too small to make much of a practical difference. This estimate implies that the $0.70 per year rise in the US federal minimum wage experienced over the period 2007 to 2009 caused less than a 0.5 percent decline in teen employment in each of these years.

For the purpose of distinguishing statistical from practical significance, Cohen (1988) offers the following well-known rough, yet plausible, guidelines: 0.2σ for a small effect, 0.5σ for a medium effect, and anything larger than 0.8σ is a large effect. By this criterion, there is a small adverse employment effect on teenage employment from raising the minimum wage.[12] However, the magnitude of the effect size will depend on the subject matter. Hence, Cohen's guidelines might be modified according to the field and potential policy intervention investigated (see Welkowitz *et al*., 1982).

3.4 Testing for heterogeneity

When there is important heterogeneity, any measure of average effect size will not capture the true nature of the economic phenomenon in question. Because economics is largely a non-experimental science, where modeling and method choices have a large effect on reported outcomes, heterogeneity is always a serious issue. The conventional meta-approach is to test explicitly for heterogeneity. The standard, widely accepted, test for heterogeneity is Cochran's Q-test, which has a chi-squared distribution with degrees of freedom $L - 1$, one fewer than there are estimates being summarized. See standard references such as Cooper and Hedges (1994), Sutton *et al*. (2000) and Borenstein *et al*. (2009) for details of the complex formula needed to calculate the Q-test. However, a much simpler method, one that is more natural for econometricians, is available to calculate this Q-test and to test for excess heterogeneity. When a simple MRA is run with t-values (dependent variable) on precision, $1/SE$, with no intercept, the sum of squared errors is the calculated Q-test and is distributed as a chi-squared with $L - 1$ degrees of freedom.

This leads to a very logical, research-driven way to meta-analyze economic research. First, the meta-analyst begins with the naïve assumption that the reported research results are homogeneous and thereby uses simple univariate, descriptive statistics to summarize the research record. Next, this naïve assumption of homogeneity is directly tested by the Q-test. In our experience, homogeneity is *always* rejected in economics research. When heterogeneity is found, meta-analysts must attempt to explain it by using all coded moderator variables in a multiple MRA (Chapter 5). If significant heterogeneity still remains, then a random- or fixed-effects multilevel, multiple MRA should be explored to ensure that unexplained heterogeneity is not distorting previous multiple MRA results (see the schema reported in Chapter 5).[13] In this way, meta-analysis proceeds in a very structured and logical manner, dictated by the actual research record itself. Surely this is the goal of any empirical inquiry.

However, there is a statistical problem with this straightforward, logical process of choosing the proper meta-analysis model. The Q-test is widely known to have low power (Sidik and Jonkman, 2007; Sutton and Higgins, 2007); thus, finding no heterogeneity may only reflect the limitation of the test rather than the true homogeneity of the research record. As a result, a case can be made to abandon the Q-test altogether and just proceed as if there is heterogeneity in all cases. This is our advice. In our experience across several dozen meta-analyses of economics research, the Q-test *always* indicates heterogeneity, in spite of its low power. Thus, it is unlikely to matter in practice whether or not the Q-test is calculated. Regardless of the outcome of the Q-test, multiple MRAs need to be employed to explain potential heterogeneity (Chapter 5).

3.5 Recap: summarizing research

All empirical analysis should include descriptive summaries of the data used. Meta-analysis is no different. Graphs such as funnel plots and frequency distributions can be used to illustrate the distribution of reported empirical findings and should be reported routinely. In doing so, they give a vivid picture of the state of empirical knowledge in a given area of research and assist with detecting coding errors, outliers, and overly influential studies. In addition to graphs, weighted and unweighted averages can be routinely reported but cannot be relied upon due to likely distortion from publication bias and systematic heterogeneity.

Even though these descriptive statistics are quite elementary, they can impart surprising insight into an area of research. However, any investigation of simple descriptive statistics must be regarded as exploratory. The insights gleamed from these simple summaries must be confirmed by formal statistical analysis and multivariate explanatory MRA, which are the topics of the next several chapters. Nonetheless, these simple descriptive statistics can be subtle, containing important nuances for the meta-analysis of economics and business research, when interpreted with caution and insight. We recommend:

- using simple graphs to reflect the distribution of reported research;
- checking for outliers and leverage points;
- avoiding vote counting and combining *p*-values altogether;
- reporting fixed-effects and random-effects averages with caution or not at all.

Meta-analysts should note that statistical significance of the overall effect size is a necessary, but not a sufficient condition for empirical relevance or for policy importance. True empirical importance further requires that the overall estimated (and corrected) effect be large enough to have a notable economic impact.

4 Publication bias and its discontents

> [P]ublication bias is leading to a new formulation of Gresham's law – like bad money, bad research drives out good.
>
> (Bland, 1988: 450)

> The house of social science research is sadly dilapidated. It is strewn among the scree of a hundred journals and lies about in the unsightly rubble of a million dissertations.
>
> (Glass *et al.*, 1981: 11)

Understanding economic phenomena requires an unbiased assessment of the state of our scientific knowledge. Politics, ideology, and vested interests routinely distort or selectively interpret the "facts." Do we need a mere reflection of published research? Or rather, is an unbiased assessment of the underlying empirical phenomenon what we require? If the motivation of our review is to evaluate the effectiveness of some social policy or economic program, then it is the latter. Or, if we seek to assess the validity of a given economic theory, it is not sufficient to merely reflect current practice but to probe a bit deeper to see whether the theory actually holds. "[E]ven a careful review of the existing published literature will not provide an accurate overview of the body of research in an area if the literature itself reflects selection bias" (De Long and Lang, 1992: 1258).

4.1 Publication selection

Publication selection is a widely accepted fact in the social and medical sciences and a severe threat to statistical inference and scientific practice (Sterling, 1959; Tullock, 1959; Feige, 1975; Rosenthal, 1979; Glass *et al.*, 1981; Lovell, 1983; Hedges and Olkin, 1985; Begg and Berlin, 1988; De Long and Lang, 1992; Card and Krueger, 1995a; Sterling *et al.*, 1995; Copas, 1999). Publication selection is largely the process of choosing research papers, or their results, for statistical significance. As a result, larger, more significant, effects will be overrepresented in the research record.

Card and Krueger (1995a: 239) identified three sources of publication selection in economics:

1 Reviewers and editors may be predisposed to accept papers consistent with the conventional view.
2 Researchers may use the presence of a conventionally expected result as a model selection test.
3 Everyone may possess a predisposition to treat “statistically significant” results more favorably.

When the majority of reported findings are selected for statistical significance, empirical phenomena can be manufactured. For example, the efficacy of a particular pharmacological treatment or the adverse employment effect of raising the minimum wage is seen by many researchers as established fact, yet these effects may be nothing more than the outcome of publication selection (Krakovsky, 2004; Doucouliagos and Stanley, 2009).

“It is a fact of life that people polish their goods to make them as shiny as possible to attract customers” (Doucouliagos and Paldam, 2009: 445). In reviewing the effectiveness of development aid, Doucouliagos and Paldam (2009) identify a “reluctance” on the part of researchers to go against the prevailing view. “To find a negative effect of aid is to question this ‘do-good’ enterprise; hence the ‘reluctance’” which they argue arises out researchers’ priors “to be seen as ‘good’, and their activity to have a ‘good’ purpose” (Doucouliagos and Paldam, 2009: 445). Most researchers and reviewers wish to make a positive contribution to the laudable enterprise of development aid. Publication bias need not arise from any nefarious motive. Rather, it is often the unintended consequence of good intentions or sound scientific practices. That is, publication selection is likely to be unavoidable – all the more reason to be aware of it and to correct its adverse effects.

The real problem of publication selection is not its existence, but the large biases that it can impart upon any summary of empirical economic knowledge, when uncorrected. For example, the average reported value of a statistical life (VSL) is likely to be biased by a factor of 5 or more (Doucouliagos *et al*., 2012b), and the adverse employment effect of minimum wage is exaggerated manyfold (Doucouliagos and Stanley, 2009). Doucouliagos and Stanley (2012) document how publication selection may represent a serious problem (“substantial” or “severe” publication selection) in nearly two-thirds of the empirical areas of economics.

Publication selection has also been found to be widespread within other sciences: the natural sciences (Sterling *et al*., 1995), political science (Gerber *et al*., 2001; Gerber and Malhorta, 2008), and medical research (Hopewell *et al*., 2009). After the widely publicized discoveries that Paxil and Vioxx have known, but unreported, life-threatening side effects, the best medical journals changed their publication policies to require the prior registration of all clinical trials (Krakovsky, 2004). When ignored, publication selection can distort any literature review, whether it is a conventional narrative review or a meta-analysis (Laird and Mosteller, 1988; Stanley, 2001). Systematic reviews of medical treatments now routinely use funnel graphs to discuss potential publication bias (see the *Cochrane Reviews*).

Box 4.1 A biased misnomer?

Publication selection bias is somewhat of a misnomer because editors and reviewers need not actively select papers or their findings to produce this bias. Often the authors themselves will report only those findings they believe to be 'correct', more rigorous, or more likely to be published at some later date. It would be more descriptively accurate to call this problem "selective reporting bias."

As we discuss in the next chapter, heterogeneity may also cloud and distort research. However, the effects of heterogeneity are usually less one-sided, less biased, than publication selection. It is the selection of random misspecification biases, heterogeneity, and sampling error that generates publication bias.

4.2 Funneling research to identify and correct publication selection bias

> The simplest and most commonly used method to detect publication selection is an informal examination of a funnel plot.
>
> (Sutton *et al.*, 2000: 1574).

A funnel graph is a scatter diagram of all empirical estimates of a given phenomenon and these estimates' precisions (i.e. the inverse of the estimates' standard errors, 1/*SE*). A clear example of the expected funnel shape can be seen among econometric studies of the productivity effects attributed to unionization (Figure 3.1, repeated below as Figure 4.1). Because a measure of the variability

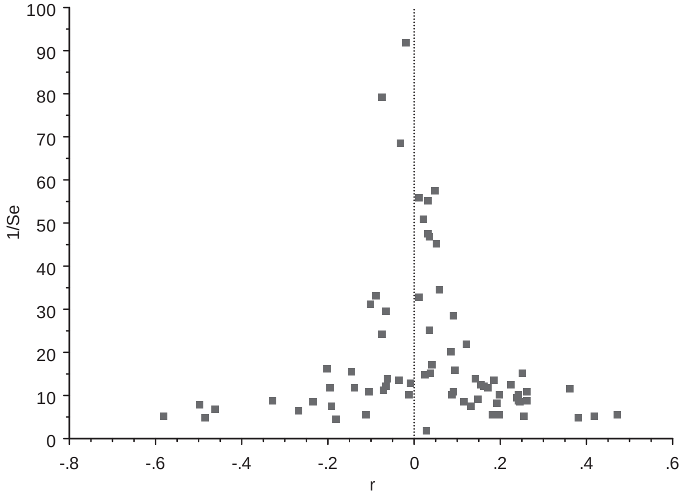


Figure 4.1 Funnel plot of union-productivity partial correlations

Source: Doucouliagos *et al*. (2005).

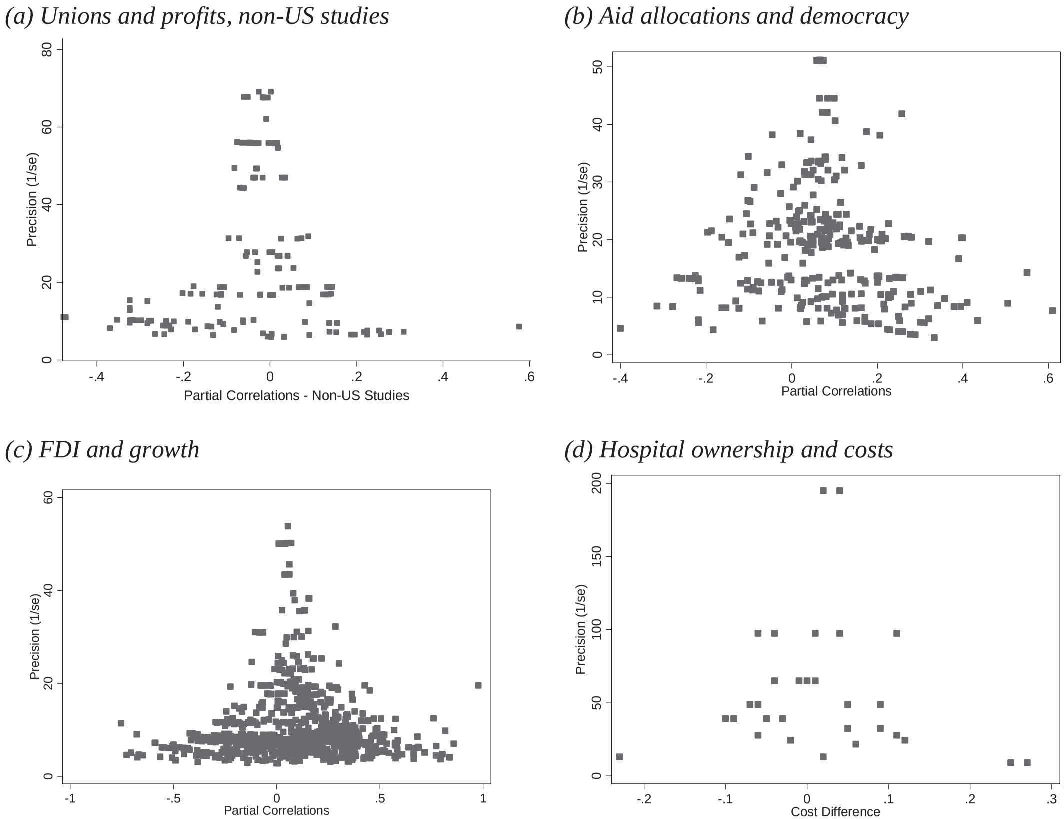


Figure 4.2 Symmetric funnel plots

of each estimate (1/*SE*) is placed on the vertical axis, those estimates at the bottom have larger standard errors and will, therefore, be widely dispersed. In contrast, the more precise estimates (i.e. those at the top) will be more compactly distributed. The union-productivity literature (Figure 4.1) provides a rough approximation to the expected inverted funnel shape that the reviewer should expect to see when there is no publication selection. Unfortunately, such approximately symmetric funnel plots are the exception. Nonetheless, a few other areas of economics research have more or less symmetric funnel graphs (see Figures 4.2).[1]

More typical is Figure 4.3, which plots the reported price elasticities for residential water demand (Dalhuisen *et al*., 2003). Figure 4.3 shows an elongated left tail with a largely missing right-hand side. Researchers who find a positive price elasticity will be unlikely to report it, thinking that it must be in error (i.e. number 2 on Card and Krueger's list of sources of publication selection, in Section 4.1 above). The asymmetry of the funnel graph is the antecedent of bias; therefore, funnel asymmetry is the key to identifying when a given area of research suffers from publication selection.

Clearly Figure 4.3 is asymmetric, reflecting publication bias and that negative price elasticities are preferentially reported. But how large is this bias and will it make any practical difference? Here, the average reported price elasticity of residential water demand is −0.38; whereas the top of the funnel is about −0.1.

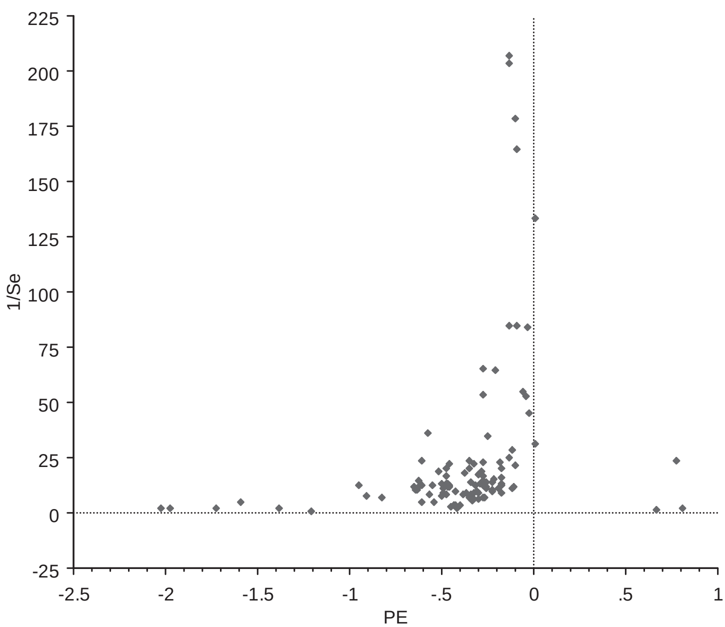


Figure 4.3 Funnel graph of price elasticity for water demand

Source: Dalhuisen *et al*. (2003).

Box 4.2 Selection paradox

Many economists have difficulty seeing that the suppression of positive price elasticities is somehow "bias." After all, we all know that raising the price of water will not increase its consumption. Surely an estimate of price elasticity that is positive must be in error. Random sampling errors will occasionally cause an estimate to be positive even when price elasticity is negative, and the likelihood of such a "bad" estimate increases for small samples, noisy data, and misspecified models of demand. This likelihood increases, the smaller is the effect. Thus, would not the economist who finds a positive price elasticity, but fails to report it, only be doing economic science a favor?

Publication selection may result from the best of motives. It is individually rational, or at least defensible, for economists to suppress positive price elasticity estimates, especially if they have any reason to suspect their data. Doing so will often improve the accuracy of the alternative negative price elasticity that the researcher chooses to report.

However, such behavior can lead to an interesting *paradox* and another economic example of the *fallacy of composition*. When the entire community of researchers suppresses positive price coefficients, the average reported elasticity, however calculated, will be biased and much larger than true price responsiveness. Although it is possible that each resulting estimate is improved when a positive estimate of price elasticity goes unreported, our collective understanding of price responsiveness worsens.

The most accurate, or precise, estimates are at the top of a funnel graph. These estimates will be least affected by publication selection because their high precision make them less likely to be statistically insignificant. But how can we identify the "top" of a funnel graph? Simulations show that averaging 10 percent of the most precise reported estimates goes a long way towards correcting publication bias (Stanley *et al*., 2010). For water elasticities, the average of the top 10 percent is −0.105, −0.106 for the top four, and the most precise price elasticity is −0.122. Publication selection bias distorts price responsiveness by a factor of three to four. The manager of the local Water Board who doubles the water rate, seeking a 38 percent conservation of water, will be quite disappointed to find that this has little effect on water use.

Our third focused example involves the value of a statistical life. These values make no claim to quantify the multifaceted joys, meanings and tragedies of the human condition. Rather, researchers observe people's behavior as they engage in voluntary risky behavior, such as choosing occupations, purchasing extra safety devices, and buying insurance. Such behaviors allow economists to impute the value that people are placing on their own lives. Needless to say, such a VSL can be controversial but is also essential for the planning of many governmental programs such as reducing toxins in our environment, improving the safety of transportation, or the construction of public infrastructure. So useful, in fact, that there have been 14 meta-analyses on the subject.[2]

Box 4.3 Symmetry exceptions

The symmetry of a funnel graph follows from the symmetry of the statistical estimates being graphed. When researchers report *t*-values of their estimates, they have already assumed that these estimates are independent of their standard errors. Otherwise, the *t*-statistics would not be valid. Nonetheless, there are exceptions to this idea that estimates will be independent of their standard errors and symmetrically distributed around the true population parameter. If a non-linear transformation of statistical estimates is used, symmetry is no longer guaranteed. For example, this can happen for non-market environmental values that are derived from estimates of demand or large partial correlations which are non-linearly transformed from regression coefficients (Stanley and Rosenberger, 2009; Stanley and Doucouliagos, 2010). Another exception is the AR(1) coefficient for a non-stationary time series, which is well known to have a non-standard and skewed distribution.

Figure 4.4 displays 39 VSL estimates, in millions of US dollars (2000 base year), calculated from the coefficients of a variable that represents the probability of death in a hedonic wage equation. Clearly, these values are highly skewed to the right, indicative of publication bias.[3] Note that there are no negative VSLs. It appears that researchers use a positive VSL or, equivalently, a positive coefficient on the probability of death as a model selection criterion. Negative values are just not economically plausible – recall number 2 of Card and Krueger's sources of publication bias.

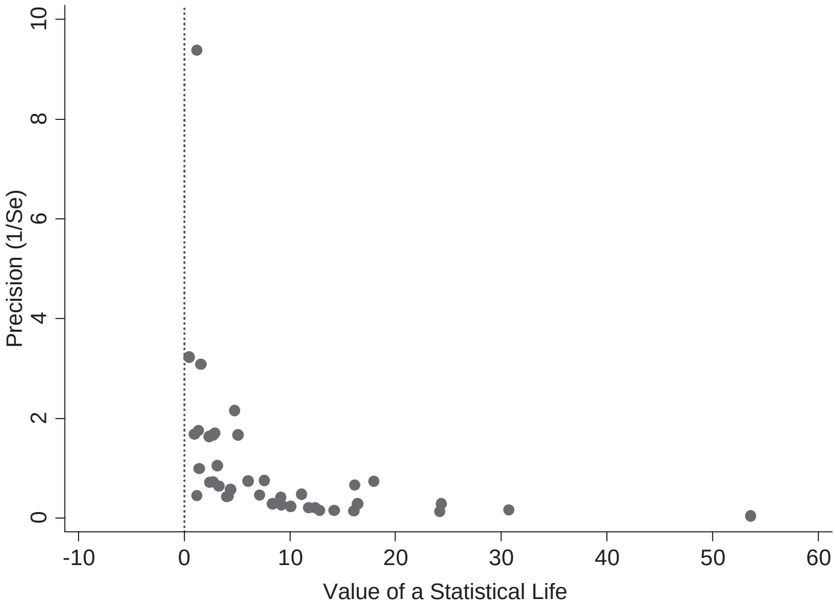
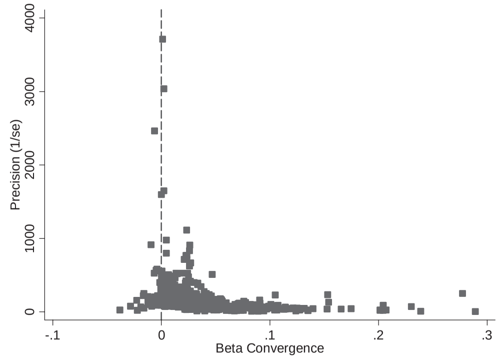


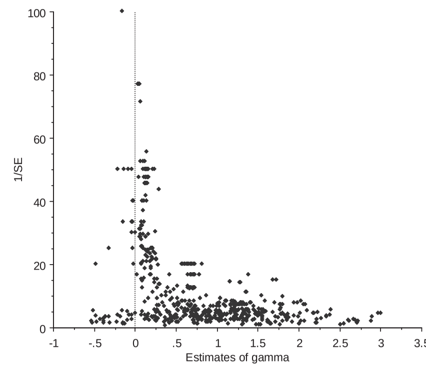
Figure 4.4 Value of a statistical life (in millions of 2000 US dollars)

Source: Bellavance, Dionne, and Lebeau (2009).

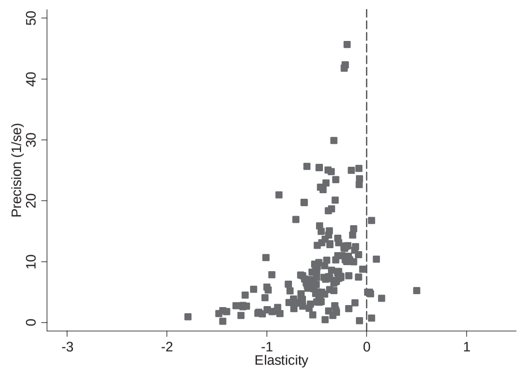
(a) Beta convergence



(b) Common currency



(c) Tobacco price elasticity



(d) Alcohol price elasticity

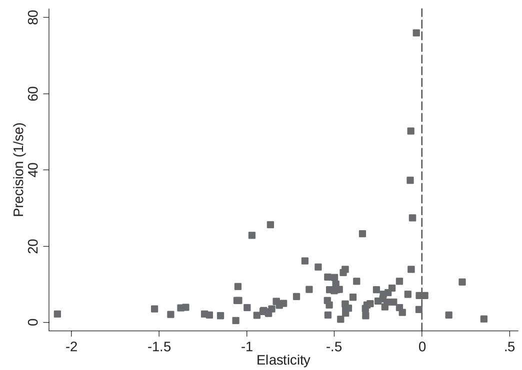


Figure 4.5 Asymmetrical funnel plots

The top of a funnel graph is less susceptible to selection bias and is therefore a better indicator of VSL. The top of Figure 4.4 is somewhat less than $2 million. The VSL as estimated by the most precise hedonic wage estimate is $1.2 million, while the average of three most precise values is $1.1 million. In any case, the top is much less than the mean of all 39 estimates, $9.5 million. Other clear examples of highly asymmetric funnel graphs are given in Figures 4.5.[4]

Our last example concerns the employment effect of the minimum wage. Recall that Card and Krueger (1995b) created quite a controversy by reporting evidence, both quasi-experimental and econometric, that minimum wage raises do not have adverse employment effects. We expand and update Card and Krueger's (1995a) meta-analysis by adding 50 studies and more than 1,400 estimated employment elasticities of US minimum wages raises (Doucouliagos and Stanley, 2009). Although the funnel of minimum-wage employment elasticities is roughly funnel-shaped, the left-hand side has many more points, especially at lower precision. This is exactly what selection for statistical significance should look like. Although positive employment elasticities are reported, they are seen less frequently (24 percent). Given the historical dominance of the competitive labour market model in economics, the preference to report significant adverse employment effects should come as no surprise. Card and Krueger's (1995a) accusation of publication bias in the minimum-wage literature seems well justified and is corroborated by objective statistical tests (see the next section).[5]

Note that the top of the minimum-wage funnel (Figure 4.6) is not much different than zero, implying that raising the minimum wage in the USA had little effect on employment. Averaging the top 148 elasticities (10 percent) gives an

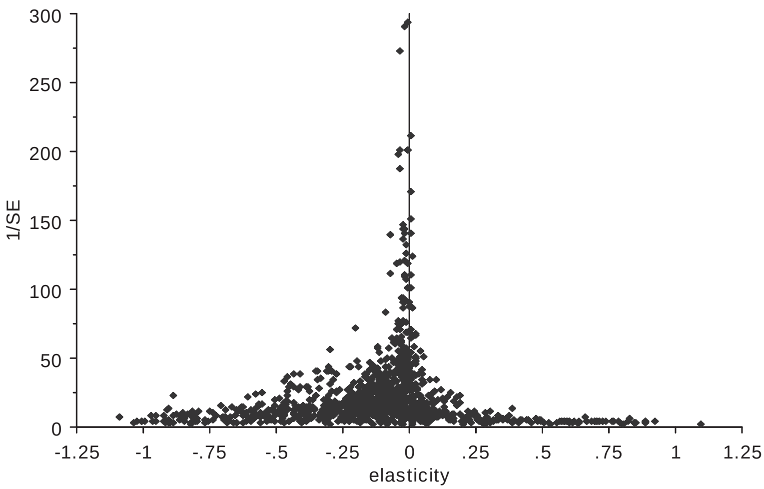


Figure 4.6 Funnel graph of estimated minimum-wage effects (n = 1,424)[6]

Source: Doucouliagos and Stanley (2009).

average of −0.02, which is not practically different than zero, the top four have an average of −0.008 (again, not significant), and the most precise estimate is negative but not significantly different from zero. Just looking carefully at this simple scatter diagram corroborates Card and Krueger's controversial finding that there is publication bias in minimum-wage research. Without publication selection, no evidence of an adverse employment effect remains. However, such issues are too important to be decided by the subjective interpretation of any graph.

"Believing is seeing" (Demsetz, 1974: 164). Thus, we need other means to assess publication selection bias more objectively. Next, we turn to objective statistical tests that corresponds to these funnel graphs. These MRA tests can identify publication selection and a genuine effect beyond publication selection should it exist.

Box 4.4 Topping out at zero?

It is important to note that the top of a funnel graph can be anywhere. It is not constrained to be around zero. In the examples that we selected, it is just a coincidence that their tops seem close to zero. The corrected price elasticity of water demand is definitely not zero; however, its top seems near zero (−0.1; see Figure 4.3), likewise for the VSL. In the absence of publication selection bias, estimates should be randomly and symmetrically distributed around the true population parameter, whatever its value.

4.3 Simple meta-regression models of publication selection

> "[T]esting of hypothesis" is frequently merely a euphemism for obtaining plausible numbers to provide ceremonial adequacy for a theory chosen and defended on a priori grounds.
>
> (Johnson, 1975: 92)

With publication selection, researchers who have small samples and low precision will be forced to search more intensely across model specifications, data, and econometric techniques until they find larger estimates. Otherwise, their results will not be statistically significant. In contrast, researchers with larger studies need not search so hard from the practically infinite model specifications to find statistical significance and will thereby be satisfied with smaller estimated effects.[7] When publication selection is present, the reported effect is positively correlated with its standard error, *ceteris paribus*; otherwise, estimates and their standard errors will be independent, as required by the conventional *t*-test and guaranteed by random sampling theory.

Such considerations suggest that the magnitude of the reported estimate will depend on its standard error, or

$$effect_i = \beta_0 + \beta_1 SE_i + \varepsilon_i \tag{4.1}$$

where $effect_i$ is an individual estimate and SE_i is its standard error.[8] $\beta_1 SE_i$ models publication selection bias, and estimates of β_0 serve as corrections for publication bias (as $SE_i \to 0$, $E(effect_i) \to \beta_0$). However, simulations have shown that it is somewhat better to use the variance, SE_i^2, in equation (4.1) rather than the standard error to estimate the genuine effect, corrected for publication bias.[9]

The error term, ε_i, in equation (4.1) is not expected to be independently and identically distributed. When $effect_i$ is an estimated regression coefficient from a large sample, it will be approximately normal and independent of other estimates. See Chapter 6 for a more detailed theoretical discussion of how the structure of MRA is derived from econometric theory and the assumptions made in the research papers that provide the values for $effect_i$. However, we know that the variance of $effect_i$, and hence ε_i as well, will typically vary from one estimate to the next. Thus, meta-regression model (4.1) has obvious heteroskedasticity and should never be estimated by ordinary least squares (OLS). Recall that SE_i is the standard error of the estimated effect, the dependent variable in equation (4.1); thus, $effect_i$ has different estimated variances, typically very much so. In practice, the differences among the reported variances are often several orders of magnitude. Consequently, weighted least squares (WLS) is routinely employed. Most statistical software calculates the WLS version of (4.1) by weighting the squared errors with the inverse of each estimates' variance (i.e. $1/SE_i^2$). Equivalently, we can divide equation (4.1) through by SE_i:

$$t_i = \beta_1 + \beta_0(1/SE_i) + v_i \tag{4.2}$$

where t_i is the t-statistic of each individual estimated empirical effect, $1/SE_i$ is its precision, and $v_i = \varepsilon_i/SE_i$, which should make its variance approximately constant. When we begin with the variance in equation (4.1), WLS becomes

$$t_i = \beta_1 SE_i + \beta_0(1/SE_i) + v_i \tag{4.3}$$

Note that there is no intercept in this meta-regression model.[10] Estimates of β_0 from either equation (4.2) or (4.3) have been shown to be among the best in comprehensive simulations of alternative corrections for publication bias (Stanley, 2008; Stanley and Doucouliagos, 2007, 2011; Moreno *et al.*, 2009a). Moreno *et al*. (2009a) call these estimators Egger and Egger var, respectively, even though "Egger var" is not found in Egger *et al*. (1997). The idea of using the variance rather than the standard error in equation (4.1) came from a "thought experiment;" see Box 4.8 below and Stanley and Doucouliagos (2007).

4.3.1 Funnel-asymmetry testing

Table 4.1 reports the results of meta-regression model (4.2) for our example meta-datasets. First, we employ MRA to identify the presence of publication selection.

Testing H_0: $\beta_1 = 0$ serves as a test of whether or not there is publication selection. This test may be considered as a test of whether the funnel graph is asymmetric, hence it is called the funnel-asymmetry test (FAT).

Box 4.5 FAT, graphically derived

To visualize how the MRA coefficient, β_1, could represent funnel asymmetry, first "invert" a funnel by plotting SE_i, rather than precision, on the vertical axis. Next, reverse the axes by placing the estimates on the vertical axis and SE_i on the horizontal. To this scatter MRA model (4.1) fits a least squares (or WLS) line. If there were too many points on the right-hand (left-hand) side of the original funnel graph, there will now be excess points on the high (low) side. To minimize the sum of the squared errors, a line will now be pulled up (down), giving it a positive (negative) slope and a positive (negative) β_1.

With the exception of the symmetric funnel graph of union-productivity research, all these intercepts are statistically different from zero. We see clear evidence that water price elasticities are skewed towards negative values (reject H_0: $\beta_1 = 0$; $t = -7.27$; $p < 0.001$), VSLs are selected to be positive (reject H_0: $\beta_1 = 0$; t = 6.67; $p < 0.001$), and adverse minimum-wage employment effects are preferentially reported (reject H0: $\beta_1 = 0$; $t = -4.49$; $p < 0.001$). For minimum-wage employment effects, water price elasticities and the VSL, we now have objective evidence that there is publication selection bias, confirming our previous visual impressions of these funnel graphs (Figures 4.2, 4.4 and 4.6). What appears to the eye to be asymmetric and skewed is confirmed by statistical tests. But is there any statistical evidence for these conventional economic effects once we make due allowance for publication selection bias? To answer this question, we turn to the precision-effect test.

4.3.2 Precision-effect testing

Next, notice the coefficients on $1/SE_i$ in Table 4.1. Testing H_0: $\beta_0 = 0$ serves as a test of whether or not there is genuine underlying empirical effect beyond the potential distortion due to publication selection (Stanley, 2005a, 2008). Because β_0 is the

Table 4.1 Simple meta-regression analysis of publication selection (dependent variable = *t*)

Variables	*Union-productivity*	*Water elasticity*	*Statistical life*	*Minimum wage*
Intercept: $\hat{\beta}_1$	0.65 (1.72)*	−2.86 (−7.27)	3.20 (6.67)	−1.60 (−4.49)
$1/SE_i$: $\hat{\beta}_0$	−0.0179 (−1.06)	−0.0817 (−5.34)	0.808 (3.56)	−0.0094 (−1.09)
n	73	110	39	1,474

* *t*-values reported in parentheses are from heteroskedastic-robust standard errors.

coefficient on precision in equation (4.2), this test is called the precision-effect test (PET). Notice further that only the price elasticity of water demand and the value of a statistical life have statistically significant $\hat{\beta}_0$s (reject H_0: $\beta_0 = 0$; $t = \{-5.34; 3.56\}$; p <.001); see Table 4.1. Even though these two research literatures are the most asymmetric and skewed thereby imparting the largest relative amount of publication bias ($\hat{\beta}_1 = \{-2.86; 3.20\}$),[11] we can still see through this fog of preferential selection to identify a genuinely negative price effect on residential water consumption and an authentic positive VSL. That is, we have reason to believe that raising the price of water will, in fact, reduce water consumption, but not by very much, and that workers do need to be compensated in the form of higher wages to take, voluntarily, higher risks. For our other examples, we find no statistical evidence of any adverse minimum-wage effect (accept H_0: $\beta_0 = 0$; $t = -1.09$; p >.05) nor evidence of any productivity effect from union membership (accept H_0: $\beta_0 = 0$; $t = -1.06$; p >.05). Thus, only water price elasticities and the VSL pass the PET.

Once we allow for publication selection bias, what are the overall empirical effects in these important areas of economics research? Often, it is the magnitude of the empirical effect, say an elasticity, that embodies many the important economic questions. The estimate of β_0 is such a corrected estimate of empirical effect, but this estimator has its problems. When there is no effect, it is biased upward, in magnitude, and when there is an effect, its bias is downward (Stanley, 2008). Consequently, we are better off just assuming that the effect is zero if a research area fails to pass the PET (i.e. accept H_0: $\beta_0 = 0$).

Box 4.6 Science is not democratic

The majority should not rule in science. Many reviews count the number of studies or results that are positive (or significantly positive), negative (or significantly negative) and insignificant. This practice is seriously flawed in even the best of cases. Hedges and Olkin (1985) show that the probability of a majority count coming to the wrong conclusion *increases* as more research accumulates. With the possibility of publication bias, majority (or plurality) rule will often come to the wrong assessment of a scientific field of inquiry. The vast majority, 76 percent, of reported minimum-wage elasticities are negative (46 percent significantly so), while only 7 percent are significantly positive. Conventional reviewers come down on the side of these negative employment effects, even though a systematic review that acknowledges known publication selection finds no adverse employment effect (Doucouliagos and Stanley, 2009). As in politics, the majority is easily manipulated.

In our examples, VSL and water elasticities pass the PET. The precision coefficients are only \$0.808 million and −0.082, which are 8.5 percent and 22 percent of the unweighted average VSL and price elasticity, respectively. In other words, 78 percent of the average reported price responsiveness is publication bias, and −0.082 is very inelastic demand. To achieve, say, a 50 percent reduction in residential water consumption, our corrected elasticity implies that prices would need to be raised by 610 percent, which is a lot more than the 136 percent price increase implied by the average elasticity. The point is that publication selection can make a huge *practical* difference to even our most widely accepted economic phenomena. This difference is even larger for the VSL.

As well as being statistically insignificant, the other two areas of economics have *practically* insignificant precision coefficients as well. Take, for example, the corrected estimate of the minimum-wage elasticity (−0.009). This implies that a doubling of the minimum wage would cause less than 1 percent of employed teenagers to lose their jobs.[12] Even if this effect were statistically significant, it is negligible from any practical policy perspective.[13]

4.3.3 Limitations

Thus far, we have discussed how publication selection can be identified and how the presence of a genuine effect, robust to publication selection bias, can be tested using meta-regression analysis. However, these tests do have their weaknesses. The FAT, which identifies the presence of publication selection, is known to have low power (Egger *et al*., 1997; Stanley, 2008). The PET, which identifies whether there is actually an empirical effect beyond the bias of publication selection, is usually powerful enough, but it can have inflated type I errors and mistakenly detect effects that are not there (Stanley, 2008). These inflated type I errors occur when there is much excess unexplained heterogeneity in the meta-regression model.

Box 4.7 Statistical software

Meta-analysis is not software dependent. Any statistical software package will be just fine for funnels, a simple scatter diagram, and for all MRA models, which are basic linear regressions. However, some of the canned meta-analysis routines should be avoided. In particular, the **metafunnel** command in STATA can be counter-productive and potentially misleading. True, it automatically produces a funnel graph, of sorts, with the 95 percent confidence limits indicated. However, the main problem is that it uses the fixed-effect estimator (FEE) to establish its center, and we know this estimator is highly biased when there is publication selection (Stanley, 2008). The whole point of a funnel graph is to get a visual sense whether it is symmetric and its approximate top. But both of these funnel functions are easily perverted when the eye is drawn to the wrong place (REE is worse than FEE). Another canned STATA routine is **metareg**. It is a weighted regression that contains a random-effects component. Because the standard error, or precision, is always one of the independent variables in our MRA models, a random-effects model is likely to be invalid. In order for a random-effect regression to be appropriate, the random components need to be independent of all of the independent variables. In economics and business applications, this is unlikely to be the case. We recommend that meta-analysts do not use these canned meta-routines but rather basic regression routines and scatter graphs.

Simulations show that PET is reliable unless there is strong evidence ($p < 0.001$) that the majority of the MRA error variance is unexplained heterogeneity (reject H_0: $\sigma_e^2 < 2$) (Stanley, 2008). When detected (i.e. if we reject H_0: $\sigma_e^2 < 2$; $p < 0.001$), we should not rely on these simple MRA models of publication selection alone but rather use a "multivariate" MRA to explain systematic heterogeneity. Such "multiple" MRAs (multiple in the sense that more than one independent variable is used) are routinely reported in economics and are expected to be part of any competent meta-analysis, regardless of any auxiliary test. These multiple MRAs are the subject of the next chapter.

However, it should be pointed out that these limitations do not affect our assessments of the four example economics literatures, because the weaknesses of our MRA methods favor the opposite of what we found. Even with the lack of power for the FAT, we found significant publication bias in three of the four examples. If anything, this limitation would suggest that the remaining area of research, union-productivity, might have unidentified publication selection. However, whether it does or does not is not really scientifically important. The funnel graph shows publication bias is, if anything, in both directions (Stanley, 2005a). Approximately balanced selection gives the meta-analyst little concern, because it has only a small effect on summary measures of overall empirical effect. The corrected effect of unionization on productivity (−0.02) is practically negligible. So what would it matter for our current assessment that unionism has no economically meaningful productivity effect if there were also undetected publication bias, in one or both directions?[14]

The PET's Achilles heel is that it can find a genuine effect too often. Only water demand and VSL have potential type I errors, because only they pass the PET (i.e. reject H_0: $\beta_0 = 0$). But who would wish to deny either that there is some positive value to life or some, albeit *small*, price effect on residential water consumption? The great thing about our meta-analysis is that magnitudes of these effects are much reduced, thereby permitting decision makers to avoid a disappointing water conservation policy or unnecessarily high protection costs. These MRA methods for publication selection bias and its correction are not perfect or "bullet-proof." However, for all of the areas of economics that we investigate here, their potential limitations only make our findings more conservative.[15]

4.3.4 PEESE: Correcting publication selection bias

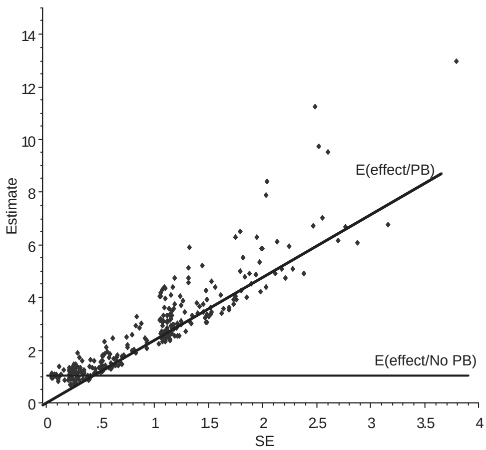
As discussed above, the coefficient on precision in equation (4.2) gives a biased estimate of empirical effect when there is publication selection, but then so do all other approaches. Although simulations have shown that this estimator is often an improvement and among the least biased estimators (Stanley and Doucouliagos, 2007; Moreno *et al*., 2009a), we still wish to do better. Stanley and Doucouliagos (2007, 2011) offer an improved correction for publication selection that uses the variance (i.e. the square of the standard error) in MRA model (4.1). Recall equation (4.3). This estimator has been dubbed the *precision-effect estimate with standard error* (PEESE), due to the form of its WLS-MRA model (4.3). The PEESE is the

MRA coefficient on precision ($1/SE_i$) from MRA model (4.3). Recall that there is no intercept in (4.3). Meta-analysts may also use the estimated intercept, $\hat{\beta}_0$, from the following equation if they use a WLS routine with $1/SE_i^2$ as the weights:

$$effect_i = \beta_0 + \beta_1 SE_i^2 + \varepsilon_i \tag{4.4}$$

Box 4.8 A thought experiment

There is a more intuitive reason why the PEESE parabola will fit the data better than a line. If there were no selection for statistical significance, then reported estimates will vary randomly around β_0, regardless of the standard error, represented by a horizontal line: E(*effect* | No PB) below. On the other hand, if there were no actual effect (β_0 = 0) but every reported estimate were statistically significant, then the expected effect would be a little larger than 2*SE*, regardless of *SE*, represented by the ray from the origin; see E(*effect* | PB) below. For very precise studies when there is actually an empirical effect, the expected effect will be many times its standard error without publication selection, and the horizontal line will dominate. But as SE gets larger, the demand for statistical significance will gradually become more dominant, and the ray, E(*effect* | PB), will exert increasing attraction upon the reported effects. It is this gradual dominance of publication selection that allows a parabola (i.e. SE_i^2) to approximate the relationship between reported effects and their standard errors.



Simulations show that the PEESE provides a better estimate of the underlying "true" effect when there is an effect (Stanley and Doucouliagos, 2007, 2011; Moreno *et al.*, 2009a). However, this is not true when there is no empirical effect and only publication selection. When the true effect or population parameter is zero (i.e. α_1 = 0),[16] we can show that the linear MRA model (4.1) is correctly specified, and its WLS estimate of the precision coefficient in MRA model (4.2) is less biased (see Section 6.3). Nonetheless, when there is no effect (i.e. α_1 = 0), but publication selection, both MRA corrected estimators are biased upward. To be conservative, we recommend that the PEESE corrected estimate of $\hat{\beta}_0$ from (4.3) be used only if we have reason to believe that there is a non-zero effect (i.e. rejecting H_0: β_0 = 0 using (4.2) and thereby passing the

PET).[17] See the flowchart (Figure 4.7) at the end of this chapter for a summary and visual representation of how these meta-regression methods are interrelated and should be employed.[18]

In spite of our own advice, we report the PEESE estimates for all four of our example areas of economics research (see Table 4.2). Recall that only two of these areas of research have robust evidence of a genuine empirical effect. The PEESE estimate of water price elasticity is −0.115, which as expected is somewhat larger, in magnitude, than the precision coefficient, −0.082, for MRA model (4.2). Still, the PEESE correction for publication bias lowers the simple average elasticity by 70 percent, and there is little practical difference between these two corrected estimates. Our visual estimate of the top of the funnel graph for water elasticities (−0.1) is well within the confidence intervals of both corrected estimates. Using the PEESE confidence interval (−0.086, −0.145) seems quite appropriate for this application.

Turning to VSL, the PEESE estimates the value of a statistical life to be $1.67 million, which is also consistent with a visual inspection of the top of the funnel graph (Figure 4.4). This corrected estimate is merely 18 percent of the unweighted average VSL. Needless to say, reducing the VSL more than fivefold is likely to have practical consequences for many areas of public policy. As a result, some programs that seek to reduce environmental or safety hazards will no longer be cost-effective.

In the case of union-productivity effects, the PEESE is virtually zero; thus, there is no conflict with our previous evaluations. The only potential exception is the minimum-wage literature, which fail to show evidence of a genuine adverse employment effect using robust standard errors. As expected, the PEESE is somewhat larger (−0.036) than the PET coefficient (−0.0094). Admittedly, this corrected effect, −0.036, is somewhat closer to being practically significant. On the other hand, this estimate implies that the $0.70 rise in the US federal minimum wage over the last three years (2007–09) caused less than a 0.5 percent decline in teen employment for each of these years. Note also that this corrected estimate of the employment elasticity of minimum wage is only one-fifth the average reported estimate. No matter how we measure it, there is a lot of publication bias in the minimum-wage literature. Once one accounts for likely misspecification biases, our best evidence indicates that there is no practically meaningful adverse minimum-wage effect (Doucouliagos and Stanley, 2009).

Table 4.2 PEESE estimates of corrected effect – MRA (4.3) (dependent variable t)

	Union-productivity	*Water elasticity*	*Statistical life*	*Minimum wage*
SE_i: $\hat{\beta}_1$	2.14 (1.00)*	−0.917 (−0.22)	0.325 (2.81)	−0.857 (−4.58)
$1/SE_i$: $\hat{\beta}_0$	−0.0034 (−0.24)	−0.115 (−7.76)	1.665 (5.50)	−0.036 (−10.11)
n	73	110	39	1,474

**t*-values reported in parentheses are from heteroskedastic-robust standard errors.

4.3.5 Dependence

Thus far, we have discussed simple OLS MRA models of publication selection.[19] It is our view that simple methods are often more robust and resilient to random data problems and misspecification biases than more sophisticated maximum likelihood methods. Nonetheless, when we have specific empirical evidence of a threat to validity, it is necessary to use those methods that explicitly address this threat, if for no other reason than to ensure that the simple OLS MRA's central findings are, in fact, robust. In this section, we consider the effects of dependence among the reported estimates. In the next chapter, we discuss explicitly the issues of heterogeneity and how to model systematic and random heterogeneity. Suffice it to say that although heterogeneity needs to be explored, it does not necessarily invalidate the results of these simple MRA models (see Section 6.4.1).

A common assumption that underpins all regression analysis is that the data are independent or, more technically, that the errors terms are independently and identically distributed. Violations of this assumption are common in conventional econometric applications (e.g. autocorrelation). Meta-analysts have long acknowledged the potential dependence among reported research estimates and have sought methods to accommodate dependence (Stanley and Jarrell, 1989; Stanley, 2001; Florax, 2002). Although there are several potential sources of such dependence, this issue is especially acute when multiple estimates from the same study are coded. There is always a possibility that the estimates reported in a given study share some common effect (perhaps due to the researchers' idiosyncratic choices of data or methods) missed by the meta-analyst (perhaps even unreported in the study) and thus omitted from the MRA.

Box 4.9 MRA autocorrelation?

More conventional autocorrelation is also possible in MRA. However, time trends and adjustment lags are more likely to be seen in macroeconomic data than meta-data. Nonetheless, dependence over time can occur in meta-data. For example, if an area of economics research is very contentious, recently published findings that support one side of a debate may stimulate supporters from the other side to conduct new research to bolster their side. When the data are arranged chronologically, this pattern of research may be seen as negative autocorrelation. Doucouliagos *et al*. (2005) find evidence of such a research pattern among estimates of union-productivity effects (recall Figure 4.1). Meta-analysts concerned about autocorrelation may test and accommodate it in the usual ways. In general, meta-regression analysis can use the full arsenal of econometric techniques and methods.

When multiple estimates from the same study are collected, they can be averaged across each study, eliminating the issue of dependence (Stanley, 2001). Doing so, however, reduces the degrees of freedom available to the MRA and its statistical power. Furthermore, some of the multiple estimates may be essential in statistically identifying the effect of a specific important research dimension. To account explicitly for potential within-study dependence, unbalanced panel models can be used (Rosenberger and Loomis, 2000b; Bateman and Jones, 2003).

The unbalanced panel version of MRA (4.1) becomes

$$effect_{is} = \beta_0 + \beta_1 SE_{is} + v_s + \varepsilon_{is} \tag{4.5}$$

for the *i*th estimate in the *s*th study. v_s represents an unobserved study effect, which traditionally is assumed to be either "random" or "fixed."[20] The "fixed-effect" term can be estimated by replacing v_s with $\delta\mathbf{D}$ (where **D** is a matrix of study dummy variables).[21]

The unbalanced panel version of WLS-MRA (4.2) is

$$t_{is} = \beta_1 + \beta_0(1/SE_{is}) + \mu_s + v_{is} \tag{4.6}$$

In effect, the meta-regression model (4.6) assumes that study effects operate largely through an unobserved differential propensity to select for statistical significance. To see this, multiply (4.6) by SE_{is}. The result is entirely the same as (4.5) with the exception that study effects are now interacted with the standard errors, $\mu_s \cdot SE_{is}$. It remains a question for future research whether MRA (4.6) or the WLS version of (4.5) better reflects typical economics and business research.

We prefer to use panel models in the context of the WLS panel model (equation (4.6)) or to weight MRA (4.5) by precision squared ($1/SE^2_{is}$) using a WLS statistical package ("analytic" weights in STATA).[22] Beyond correcting for heteroscedasticity, precision serves as an indicator of quality. Weighting by precision in either of these two ways limits the influence of widely dispersed and sometimes implausible effect estimates that lie at the bottom of the funnel graph.

In our previous examples, only the minimum wage meta-analysis has the full multidimensional data structure required for panel analysis. Here, we found 64 studies containing 1,474 minimum-wage estimates and their standard errors. Studies contain anywhere from 1 to 96 estimates and average 23 per study. Columns 1 and 2 of Table 4.3 reports both the "random"- and "fixed-effects" panel (or multilevel) FAT-PET-MRA results for the minimum-wage literature using (4.6). In this application, there are no practical differences between the random-effects (REML) and fixed-effects (FEML) multilevel MRAs. The corrected elasticity estimates are virtually the same, approximately −0.01. Although now statistically significant, as we discussed previously, such a small elasticity remains, nonetheless, practically negligible.

REML critically assumes that the unobserved study effects are independent of the included independent variables, *SE* or 1/*SE*. Researchers who select findings for their statistical significance will likely experiment with econometric model specifications and techniques to achieve their goal. These efforts are expected to be correlated with *SE*, because larger standard errors require larger selected study effects to achieve statistical significance. Thus, this critical assumption will be routinely violated in economic MRA applications. As a consequence, we prefer to use "fixed-effects" panel MRA models. The researcher who wishes to test which

Table 4.3 Panel and cluster MRA of publication selection among minimum wage employment effects (dependent variable *t*)

	REML:1[a]	*FEML: 2*	*Cluster-Robust:3*	*FE-Cluster: 4*	*FE-WLS: 5*	*Average: 6*
Intercept: $\hat{\beta}_1$	−1.71 (−5.62)*	−1.59 (−20.5)*	−1.60 (−4.49)*	−1.59 (−11.8)*	−1.15 (−11.4)*	−1.25 (−2.61)*
1/SE$_i$: $\hat{\beta}_0$	−0.0099 (−4.17)	−0.0097 (−4.06)	−0.0094 (−1.09)	−0.0097 (−1.64)	−0.0072 (−1.05)	−0.0151 (−1.08)
k	64	64	64	64	64	64
n	1,474	1,474	1,474	1,474	1,474	64

*t-values are reported in parentheses.

[a] REML conventionally stands for "restricted maximum likelihood." Here, it can also stand for "random-effects multilevel" because restricted maximum likelihood may be used to estimate the random-effects multilevel MRA model (4.6).

of these multilevel models is appropriate for her meta-analysis can conduct a Hausman test to decide. In a multiple MRA context, Chapter 5, it is even more likely that the unobserved study effects will be correlated with some independent variables.

The issue of dependence concerns efficiency, rather than bias. Not correctly accommodating the proper error structure in one's MRA can cause the MRA standard errors and *t*-values to be calculated incorrectly. In such cases, the usual worry is that simple methods will be too generous in calculating these statistics and might give a false appearance of statistical significance. This is not the case in the minimum-wage literature because both REML and FEML find the MRA coefficients to be statistically significant. Furthermore, efficiency is a second-order concern compared with bias. As we have seen above, publication bias is a much greater threat to understanding a given economic phenomenon, often biasing average reported effect manyfold. Efficiency issues pale in comparison to the often overwhelming effect of publication and misspecification biases.

For a more conservative assessment of the MRA coefficients, the meta-analyst can use cluster-robust standard errors. Treating each study as a cluster and thereby allowing potential dependence among the reported estimates within each study to calculate the standard errors is another sensible way to handle potential dependence. Column 3 of Table 4.3 presents the MRA results for minimum-wage employment effects using cluster-robust standard errors. Non-robust, conventional OLS standard errors make $\hat{\beta}_0$ statistically significant ($t = -3.55$; $p < 0.01$; not reported in the tables). As expected, cluster-robust standard errors provide a more conservative assessment than OLS standard errors ($t = -1.09$; $p > 0.05$; reported in Table 4.3). Whether the standard errors are conventional or cluster-robust, the estimated MRA coefficients will be identical; thus, either approach furnishes an identical, practically insignificant, corrected estimate of the effect of minimum wage on employment. To err on the conservative side, meta-analysts should routinely use cluster-robust standard errors whenever multiple estimates are coded per study.

A potential criticism of fixed-effects panel methods is that they tend to give smaller standard errors and are thus likely to exaggerate the significance of estimated MRA coefficients. But this too is easy remedied by calculating cluster-robust standard errors within a fixed-effects panel model context. Calculating cluster-robust standard errors is a menu option in STATA, and we report these fixed-effects cluster-robust findings in column 4 of Table 4.3. Note how the signal of publication selection remains very strong, but the existence of a genuine minimum wage effect becomes questionable. When the fixed-effects WLS version of MRA (4.5) is used (column 5), the exact estimated MRA coefficients are somewhat different, but the overall results are essentially the same as cluster-robust and FE-robust (columns 3 and 4).

Lastly, this potential dependence may be accommodated by running an MRA on the average study estimate and their standard errors. That is, we can easily calculate the average of all estimates, *t*-values, and standard errors in each study and run our MRA across studies. Doing so will typically result in many fewer observations and thereby a loss of efficiency (64 vs. 1,474 for the minimum-wage literature), but it has the additional benefit of weighting each research study equally. When each estimate is coded and analyzed, studies reporting a large number of estimates (e.g. 96 vs. 1) can have an undue influence on the statistical assessment of the entire research literature. As a simple approach to potential within-study dependence and to avoid any undue weighting of a literature's findings, Stanley (2001) advocates meta-analyzing study averages. Differences in how one weights each research study can reverse the overall assessment of a research literature, for example class size effects on study achievement (Krueger, 2003; Mishel and Rothstein, 2002). We still advocate this approach because it offers a more conservative and often more realistic assessment of the MRA's statistical significance. Nonetheless, we also suggest using multiple estimates with multi-level models and cluster-robust standard errors to ensure the robustness of central MRA findings.[23]

When calculating averages for each study, researchers need to consider whether they should be using a simple average or a weighted average. Hunter and Schmidt (2004) recommend strongly that a weighted average should be constructed for each study, using sample size as weights. Precision can also be used, or some other set of weights can be used depending on the circumstances.

Our recommended menu of MRA methods may seem a little eclectic or even bewildering to the uninitiated. We view the running of multiple MRA models and approaches as a sensitivity analysis, a way to ensure the robustness of our findings. In our experience, the central findings (whether there is a genuine effect and its approximate magnitude) are robust to MRA method. Our approach is to look for common patterns across several plausible models. For those who seek the "correct" model, we are agnostic. The correct model will depend on nuances in the structure of research in the specific area under investigation. Of course, one can use a battery of econometric specification tests to see which MRA model is appropriate. However, these specification tests are subject to type I errors and tend to have low power. Because multiple specification tests are required, there will be a high probability of making some error. Nonetheless, several of these tests are discussed in the next chapter. Our approach is a pragmatic compromise. Aside from the irreducible ambiguity of econometric specification testing, referees and editors will ask the meta-analyst to perform such robustness checks because this is the accepted practice in conventional econometrics.

4.4 Alternative approaches to publication selection

This chapter has introduced and illustrated basic meta-regression methods for publication selection detection and correction that center on the standard error, its square, or its inverse, precision. Needless to say, there are alternatives to the approaches advocated here. It is our judgment that those discussed above are best suited for applications in economics. Nonetheless, at the risk of being too brief and dismissive, we review other strategies for addressing publication selection bias.

Box 4.10 Root *n* for the MRA model

There are some cases where the standard errors are either unavailable or inappropriate. For example, when demand coefficients are non-linearly transformed to non-market environmental values, these values and their standard errors will be correlated even if there is no publication selection (Stanley and Rosenberger, 2009). In such cases, the square root of the sample size serves as a proxy for precision. The standard errors, or their inverse, are more accurate and complete measures of an estimator's precision than any function of the sample size alone, because sample size does not contain information on many other factors that affect variation. Nonetheless, the square root of the sample size can serve as rough proxy for precision in our FAT-PET-MRA when the standard errors are either unavailable or inappropriate (Stanley and Rosenberger, 2009).

4.4.1 Rosenthal's failsafe N

Rosenthal (1979) is an insightful early contributor to meta-analysis and in its connection to publication selection bias. He coined the term "file drawer problem" to describe colorfully the preference for statistically significant results. When there is publication selection for statistical significance, he reasoned, many unpublished and insignificant results must be languishing in researchers' file drawers.

To address this issue, Rosenthal (1979) offers a formula for the number of unpublished and insignificant papers that would need to be contained in these "file drawers" in order to reverse an overall assessment of statistical significance given by the published papers in an area of research. More precisely, this "failsafe *N*" calculates the number of studies with a zero effect that must be hidden away to bring down the average overall effect to statistical insignificance. If this "failsafe *N*" is an implausibly large number, say thousands (or perhaps even hundreds), then the reviewer concludes that there is in fact a genuine empirical effect, regardless of publication selection.

Box 4.11 Statistical vs. Economic Significance

In a series of papers spanning two decades, McCloskey (1985) has emphasized the distinction between statistical significance and economic importance (or practical significance) (McCloskey, 1995; Ziliak and McCloskey, 2004). Economic importance entirely hinges on the magnitude of an empirical effect, not merely on its sign or statistical significance. Practical significance answers the question: How large does this empirical effect need to be before anyone notices or before it makes any meaningful difference to the lives of consumers, businesses or policy makers? The failure to recognize this distinction fully has caused debates throughout the social sciences, notably psychology (Thompson, 1996, 2004; Harlow *et al.*, 1997). This is another weakness of Rosenthal's failsafe *N*. One estimate with a large bias (perhaps through misspecification) might easily be sufficient to make the average standardized effect statistically significant, regardless of its practical significance.

Unsurprisingly, there are a number of problems with Rosenthal's approach. For example, other analysts have shown that his formula is wrong (Begg and Berlin, 1988; Iyengar and Greenhouse, 1988; Scargle, 2000). Yet, more importantly for empirical economics, the logic behind Rosenthal's failsafe *N* is flawed. It is important to realize that Rosenthal is a psychologist who thinks in terms of experiments. When the failsafe *N* is in the thousands, it is reasonable to suppose that there could not be such a large number of unpublished experiments lying around somewhere. Such a large number of experiments would take too much time and resources to go unnoticed.

However, in economics, it takes almost no time or resources to produce yet another empirical estimate. Applied econometric research is largely observational, that is, the researcher does not need to spend her time collecting or creating the data. Rather, governmental and business data are downloaded and submitted to some statistical software. While doing so, econometricians are free to choose among many different independent variables to include along with the variable of interest in their econometric models, alternative measures of the dependent phenomenon, many combinations of data or data ranges, several different functional forms of the underlying relationship, and dozens of estimation techniques and approaches. Together, millions of estimates are easily generated about most key economic parameters (Sala-i-Martin, 1997). Worse still, econometricians can program their computers to generate and select among these millions of estimates to find the "best" or most significant one. As a result, economics has no finite "failsafe *N*." A calculated "failsafe *N*" as high as a million gives no guarantee that the reported statistical significant phenomenon is anything more than publication selection.

4.4.2 Trim and fill

A widely employed strategy to correct for publication bias in medical research is "trim-and-fill" (Duval and Tweedie, 2000). It begins with a funnel graph and attempts to impute a corrected estimate by "trimming" the excess reported studies on the "preferred" side of the funnel graph and "filling" in the missing, unreported studies on the other side. The central weakness of this approach to correcting publication bias is the crucial issue of how to identify these two sides of the funnel graph. Sides of a funnel graph are defined relative to the true underlying empirical effect being investigated. That is, before we can either trim or fill in the funnel graph, we must first have an idea where the true empirical effect is located. Yet, this is exactly what we are seeking from trim and fill. How can this vicious circle be broken?

Duval and Tweedie (2000) use a random-effects weighted average of all reported estimates as the first approximation to the true underlying effect. However, it is widely known that such weighted averages are highly biased when there is publication selection (Stanley, 2008; Moreno *et al*., 2009a). From this poor starting point, they estimate the number of missing, unreported studies and trim this number from the preferred side of the funnel graph. This produces a second estimate of the true effect, and this process is repeated until convergence is reached. Typically, the successive rounds of this algorithm do produce less biased estimates. However, the central weakness of trim-and-fill is that confidence intervals of the corrected estimate often do not contain the true parameter being estimated. In comprehensive simulations conducted by medical researchers sympathetic to trim-and-fill, the meta-regression methods discussed above (PET and PEESE) are found to be superior. “In conclusion, several regression-based models for [publication bias] adjustment performed better than” trim-and-fill (Moreno *et al.* 2009a: 15).

4.4.3 Hedges' maximum likelihood, publication selection estimator

Larry Hedges is another early and influential contributor to meta-analysis. Hedges and Olkin (1985) is the classic statistical text on meta-analysis. Hedges (1992) offers a more sophisticated econometric model of the publication selection process along the lines of a Heckman correction for selection. As discussed before, the first stage of the conventional Heckman method is unavailable, because we do not observe the unreported estimates. Instead, Hedges' approach assumes that the selection process is a function of an estimate's p-value, and nothing else. In particular, the likelihood of publication is an increasing step function of the complement of a study's p-value.[24]

$$w(effect_i, \sigma_i) = \begin{cases} \omega_1 & \text{if } -\sigma_i\,\Phi^{-1}(a_1) < effect_i \leq \infty \\ \omega_j & \text{if } -\sigma_i\,\Phi^{-1}(a_j) < effect_i \leq -\sigma_i\,\Phi^{-1}(a_{j-1}) \\ \omega_k & \text{if } -\infty \leq effect_i \leq -\sigma_i\,\Phi^{-1}(a_{k-1}) \end{cases} \tag{4.7}$$

where $1 < j < k$, and $\Phi^{-1}(a_1)$ is the inverse cumulative normal (Hedges and Vevea, 1996: 304). The a_js are arbitrary cut points, such as 0.10, 0.05, 0.01 and 0.001, which are chosen *a priori*. The weights of these arbitrary cut points, a_j, can be estimated from the data. After fully parameterizing this selection model, Hedges (1992) derives the joint likelihood and uses a multivariate Newton–Raphson method to find its maximum.

Hedges' maximum likelihood, publication selection estimator (MLPSE) has been applied to several areas of economic research but with mixed success; see Ashenfelter *et al.* (1999), Florax (2002), Abreu *et al.* (2005), Nijkamp and Poot (2005), and Huang *et al.* (2009). The MLPSE has been problematic in many of these applications. For example, Florax (2002) finds that the MLPSE does not converge for estimates of the price elasticity of water demand, which is one of the examples we have been using. Worse still, Florax (2002) unearths the “awkward” implication that the probability of publishing a statistically insignificant elasticity is greater than that of publishing a statistically significant one. Likewise, Abreu *et al.* (2005) obtain implausible weights for publishing estimates of economic convergence. They find that studies with insignificant p-values between 0.05 and 0.10 are more likely to be published than statistically significant ones. We believe that such patterns of selected p-values provide evidence that Hedges' publication selection model is misspecified. The story is similar for Huang *et al.* (2009). They find that highly significant social trust and social participation effects ($p < 0.01$) are less likely to be published than marginally significant ones ($0.01 < p < 0.05$), and the corrected effect of social trust is virtually the same as the unadjusted mean even though there is strong evidence of publication selection for a positive social trust effect (Huang *et al*., 2009: 460).

We suspect that Hedges' assumed selection model misspecifies publication selection in economics. We are dubious that there are multiple steps of *p*-values that should be given different weights in economics. Selection is likely to be more complicated than any function of *p*-values alone. Rather, selection for publication in economic journals will depend on whether or not an effect is statistically significant but also on other unrelated features concerning the perceived quality of the analysis. For example, methodological innovation is often prized over mere statistical significance. Or selection may be related to the author's reputation or the novelty of the data she uses. In any case, only a portion of any economics research literature is likely to be selected for statistical significance.[25] In the simulations used in our research, we assume varying fractions (from 0 to 75 percent) of a literature are selected for statistical significance while the remaining portions are assumed to be selected for unrelated reasons (Stanley, 2008; Stanley and Doucouliagos, 2007). Multiple FAT-PET-MRAs can explicitly model a more complex publication selection process, allowing it to be affected by any observed study characteristic. These more complex MRA models of publication selection and heterogeneity are the subject of the next chapter and are discussed in detail there. Maximum likelihood methods are highly sensitive to small changes in data and model specification. Thus, in practice, the results obtained using Hedges' MLPSE are less likely to be reliable.

4.4.4 Meta-significance testing

The decisive characteristic that identifies a genuine empirical effect from random misspecification biases and publication selection is that the associated standardized effect (i.e. a *t*- or *z*-value) increases with larger samples or greater precision. Statistical power guarantees that *t*-values increase with the square root of the sample size (or precision), *ceteris paribus*. Card and Krueger (1995a) were the first to use this insight explicitly in meta-regression analysis. Recall that a *t*-value is:

$$t_i = (effect_i - \alpha_1)/SE_i \tag{4.8}$$

where α_1 is the associated "true effect" (population parameter) and is assumed to be zero as the null hypothesis. Thus, we would expect that the estimated *t*-values would increase with precision ($1/SE$), assuming of course that there is some genuine effect. Whether the effect is a regression coefficient or something simpler, basic statistics tells us that SE_i will be proportional to $1/\sqrt{n}$, further implying that the estimated *t*-value should be proportional to $\sqrt{n}$.[26] Thus, the signature of a genuine underlying empirical effect is this power trace.

However, when there is no genuine effect, estimates of α_1 will vary randomly around zero, and the *t*-value will be independent of sample size. Because the probability of the type I error is constant for all sample sizes, standardized test statistics automatically adjust for differences in sample size. Therefore, when there is no underlying empirical effect, large *t*-values will be observed rarely and randomly, regardless of *n*. Of course, this all assumes that there is no publication selection bias to dilute the effect of statistical power. Alternatively, when there is a genuine empirical effect, statistical power will cause the *t*-value to be positively associated with the square root of its sample size. This positive relationship can be expected regardless of the size of the effect (assuming that it has a non-zero effect) and irrespective of contamination from random misspecification biases. This trace of statistical power identifies a genuine empirical effect across a given research literature.

These considerations suggest the following MRA model:

$$E(\log|t_i|) = \gamma_0 + \gamma_1 \log(n_i) \tag{4.9}$$

γ_1 will be zero if there is no effect, and $\gamma_1 = 1/2$ when there is a non-zero effect. This trace of statistical power has been used as a test for a genuine effect beyond publication selection and has been called "meta-significance testing" (MST: Card and Krueger, 1995a; Stanley, 2001, 2005a).

Like all tests, MST has its limitations. When there is publication selection the relationship between the *t*-statistic and its sample size washes out. But worse, MST often has type I error inflation (Stanley, 2008). In this context, the type I error is the mistake of finding that there is a genuine empirical effect when there is really none. Because scientific inference is designed to be conservative, the type I error is the worse error to make. This problem arises from the use of the absolute value before taking the logarithm of the *t*-value. This causes both very positive and very negative values to be large and positive. When there is excess heterogeneity, ubiquitous in economic applications, such large *t*-values (positive or negative) will be found more often in large samples.

Box 4.12 What might prevent the *t* ratio from rising with sample size?

Card and Krueger (1995a: 239) erroneously use the absence of the relation between the logarithm of a study's reported *t*-value and the logarithm of its sample size (or degrees of freedom) as evidence of publication bias in minimum wage research – recall equation (4.9). However, there is a second, equally valid, cause of an insignificant estimated γ_1 in equation (4.9). Perhaps, there is simply no actual empirical effect. In the absence of an empirical effect, t-values will not rise with their sample size, regardless of whether or not there is publication selection (Stanley, 2005a). Unfortunately, this oversight has been repeated several times by economists. MST should not be used as a test for the presence of publication selection. The simple answer to Card and Krueger's rhetorical question: "What might prevent the *t* ratio from rising with sample size?" (Card and Krueger, 1995a: 239) is that the minimum wage has no employment effect (Stanley and Doucouliagos, 2009).

Due to MST's type I error inflation, we reluctantly do not recommend its use.[27] Nonetheless, this idea of searching for the trace of statistical power is central to differentiating genuine empirical phenomena from publication selection bias. Fortunately, this same fundamental statistical relationship, statistical power, is also embedded in the FAT-PET-MRA model (4.2). Note how MRA model (4.2) is a regression of *t*-values on precision. Precision is literally part of a *t*-statistic (recall equation (4.8)); thus, it must be a better measure of statistical power than the square root of the sample size. Also, MRA model (4.2) does not need to use absolute values and is thereby much less vulnerable to type I error inflation. Lastly, the FAT-PET-MRA has more power than MST (Stanley, 2008). Thus, MRA model (4.2) dominates MST.

4.5 Recap: The FAT-PET-PEESE approach to publication selection

This chapter introduces publication selection and how meta-regression analysis can accommodate and correct for its bias. We believe that the best available models of publication selection are FAT-PET-MRA and PEESE. The FAT-PET-MRA is:

$$t_i = \beta_1 + \beta_0 \,(1/SE_i) + v_i \tag{4.2}$$

where t_i is the reported estimate's *t*-value, and SE_i is its standard error. This meta-regression model (4.2) contains tests for publication bias (funnel-asymmetry test, H_0: $\beta_1 = 0$), and for the presence of a genuine effect beyond publication selection (precision-effect test, H_0: $\beta_0 = 0$).

The precision-effect estimate with standard error (PEESE),

$$t_i = \beta_1 SE_i + \beta_0 (1/SE_i) + v_i \tag{4.3}$$

provides a better estimate of the actual empirical effect corrected for publication bias, when there is one. Although there are other methods to deal with publication selection, FAT-PET-PEESE are the best available meta-analytic methods for economic and business applications.

All economic and business research literatures should be assumed to contain publication bias, regardless of one's assessment of the funnel graph. In our experience over dozens of areas of economic and business research, the clear majority possess substantial publication selection bias. Although the funnel-asymmetry test is a valid test for the presence of publication selection under typical conditions, it is widely known to have low power. Thus, the failure to find explicit evidence of publication selection is no guarantee that its bias is not a serious problem.

Regardless of the outcome of the FAT (node 1 in the schema shown in Figure 4.7), we recommend using the PET to test for the existence of a genuine effect beyond potential contamination from publication bias (node 2). Lastly, PEESE should

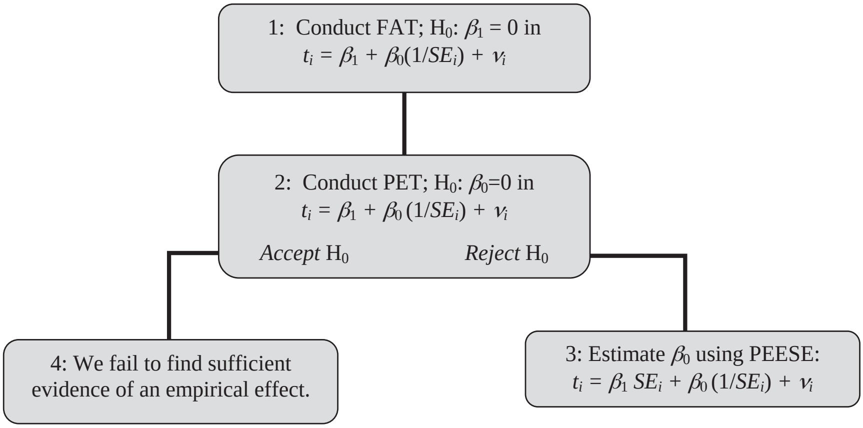


Figure 4.7 Schema for investigating and correcting publication bias

be used to estimate the magnitude of the empirical effect if there is evidence that one exists (i.e. reject H_0: $\beta_0 = 0$) – node 3. After the corrected estimate is calculated, the meta-analyst should evaluate its size for practical economic or policy significance. If there is no evidence of a non-zero empirical effect beyond publication selection (i.e. accept H_0: $\beta_0 = 0$), the meta-analyst must accept that the research literature in question has failed to provide evidence of a genuine empirical effect. In the following chapter we turn to multivariate versions of these simple MRA models.

5 Explaining economics research

Economic phenomena are complex and nuanced. Even where there is a simple underlying phenomenon, unforeseen economic and political events easily overwhelm the most stable and fundamental economic relation. For example, the added uncertainty caused by the 2008 global financial crisis broke down the normally stable relation between income and consumption expenditures and between interest rates and the demand for new loans, or at least those versions of these relationships that fail to incorporate the full effects of unobserved uncertainty.

Likewise, economics research is complex and nuanced. Even when applied econometrics is estimating a clear, meaningful and stable parameter (e.g. a price elasticity), the econometric technique employed (e.g. structural equations), the type of data (e.g. panel) or whether relevant explanatory variables are omitted (e.g. income) often has a dominating impact on the estimated coefficient. Past meta-analyses routinely find wide differences among the reported estimates of purportedly the same economic parameter. Great disparities among research finding are ubiquitous. For example, the reported minimum-wage elasticity of employment ranges from −19 to nearly +5, with a standard deviation of 1.1. Note that this variation, however measured, overwhelms the reported average elasticity, −0.19. Similarly, the price elasticity of residential water demand has an implausibly large reported range (from −2 to +0.8), implying that demand is anywhere from quite price elastic to highly inelastic or even upward-sloping! For these price elasticities, the average (−0.38) is also dominated by the reported variation (standard deviation 0.41).

In previous chapters, we have focused on describing and summarizing research findings for any given area of research. However, the most common finding among the hundreds of meta-analyses conducted on economic subjects is that there is excess heterogeneity. That is, the observed variation in any area of economics research is always much greater than what one should expect from random sampling error alone. Such excess variation begs explanation. Furthermore, the failure to account for such excess variation can invalidate any simple meta-analysis. Like conventional econometric analysis, the omission of a relevant explanatory variable might possibly cause bias in the simple meta-analysis. In this chapter, we discuss how to accommodate and explain this excess research variation using “multivariate” (or “multiple”) meta-regression analysis.

5.1 Heterogeneity

The "problem" of heterogeneity arises from the fact that the expected value of a reported estimate will often depend on many other factors: country or region, time period, presence (or absence) of other relevant variables in the original econometric model, dependent variable measure, functional form used and the econometric technique employed, among others.

If unaccounted, heterogeneity can bias any simple MRA estimate. Econometricians are well aware of omitted-variable bias. When a relevant independent variable that is correlated with an included independent variable is omitted from a regression (conventional or meta), the estimated regression coefficients will be biased. Systematic heterogeneity, not explicitly accommodated, may bias simple MRA estimates.

The most obvious approach to addressing excess heterogeneity is to explicitly model it by coding any research dimension thought to have a potential effect on the reported effects, including the standard deviation or its inverse, precision. It is standard – and highly recommended – practice to model systematic heterogeneity using a multivariate or multiple MRA.[1] Multiple MRA is the central topic for this chapter. A further advantage of meta-regression analytic explanation of research heterogeneity is that it is based on theory. This theory is statistical theory and need not be assumed by meta-analysts because the researchers of the primary empirical literature themselves must have made the necessary assumptions if their results are to be taken at face value. We discuss the theory of meta-regression analysis and the more technical aspects of the MRA in Chapter 6, while extensions to MRA, including multiple MRA equation systems, are discussed in Chapter 7. In this chapter, we first briefly evaluate the random effects approach to dealing with heterogeneity and then discuss how to use MRA to explain heterogeneity.

Box 5.1 MRA of *t* versus precision

To visualize this connection between excess heterogeneity and a simple MRA between a *t*-value and its precision, consider what homogeneity implies. With homogeneity, each estimate will be randomly distributed around its true value, β_0, $effect_i = \beta_0 + \varepsilon_i$ (recall equation (3.3)). However, we know that different estimates of effect are likely to have widely different sampling errors, SE_i, and we need to compensate for this heteroskedasticity. The simplest way to do so is to use a WLS version of (3.3), which also represents the fixed-effects weighted average. The WLS version of (3.3) may be obtained by dividing (3.3) by SEi, giving: $t_i = \hat{\beta}_0 /SE_i + v_i$, which is an MRA of *t* versus precision, $1/SE_i$. Because v_i is the former error, ε_i, divided by SE_i, it must have a variance of 1, unless there is excess heterogeneity coming from some source other than measured sampling error, SE_i, alone. Thus, the sum of squared errors from this simple of MRA of *t* vs. precision $1/ SE_i$ gives a simple chi-squared test, called the *Q*-test, for excess heterogeneity.

5.1.1 Is there excess heterogeneity in economics research?

The short answer is: yes, most definitely. In our experience, all areas of research contain excess heterogeneity; that is, greater variation than what would be expected by measured sampling error, SE_i, alone. As noted in Chapter 3, the conventional test is Cochran's Q-test, which has a chi-squared distribution with degrees of freedom one fewer than the number of estimates (Cooper and Hedges, 1994; Sutton *et al.*, 2000). The easiest way to calculate Q is to use the sum of squared errors from a simple MRA of the t-value on precision forced through the origin (Higgins and Thompson, 2002). This MRA is the same as the WLS version of the FAT-PET-MRA, except that there is no intercept.[2] Table 5.1 reports the calculated Cochran's Q for our four example meta-analyses. In all of these areas of research, there is strong evidence of excess heterogeneity ($p < 0.001$).

5.1.2 Random-effects meta-regression analysis

One approach to dealing with heterogeneity is to assume that excess heterogeneity is random and independent of all of our moderator variables, including the standard error. Such "random-effects" MRAs incorporate an additional term to the MRA model, which allows for any between-study (or between-estimate) random variation. For example, STATA has a routine, **metareg**, that includes random effects along with weighted least squares where the weights include both a measure of the between-estimate variance and the square of the standard error (the within-estimate variation).[3] Random-effects models are more familiar to econometricians in the context of cross-section/time series panel data. As discussed in the previous chapter and again below, random-effects multilevel (REML) panel models are one way to address within-study dependence.[4] Multilevel models can also be structured over different sources of data, authors or other potential data dependencies.

A crucial assumption in any random-effects model is that these added random effects need to be independent of all of the explanatory variables. However, as discussed in Section 6.2.1 below, this is not likely to be true for MRA models of publication selection because imprecise studies (i.e. those with larger standard errors) require greater effort to find statistical significance. Thus, the random effects will be routinely correlated with the standard error when there is publication selection. Random effects are then, in part, the result of these greater efforts to select and report desired estimates.

Table 5.1 Q-tests for heterogeneity (dependent variable t)

Variables	*Union productivity*	*Water elasticity*	*Value of statistical life*	*Minimum wage*
Q	347	2,533	558	14,528
df	77	109	38	1,473

Preliminary simulations using the same design as Stanley (2008) confirm a positive correlation between random, yet selected, heterogeneity and the estimate's standard error. As expected, this correlation increases with the incidence of publication selection. These simulations also suggest that random-effects MRA models discussed in Chapter 4 cause larger biases than "fixed-effects" MRA, and these differences can be of practical importance.[5] Thus, we do not recommend the use of random-effects MRAs.

To see this, recall from Chapter 4 that we found evidence of genuine non-zero empirical effects in two of the economics research examples, the price elasticity of residential water demand and the value of a statistical life. When there is evidence of such a genuine effect, the PEESE provides a better (less biased) estimate. Table 5.2 reports the previous WLS-PEESE estimates from MRA (4.4) versus the same model that employs a random effect to allow for excess heterogeneity.

Note that in both cases, the random-effects MRA greatly increases the estimated effect in the same direction as the observed publication bias; hence, we believe that the random-effects MRA increases bias. Our interpretation is based on the comparison of the "corrected" estimates in Table 5.2 with the known biased weighted averages (FEE and REE; recall Chapter 3) that are reported in Table 3.3. To see this, first recall that both of these literatures contain strong evidence of publication selection (Table 4.1). Simulations clearly reveal that both fixed-effects (FEE) and random-effects (REE) weighted averages are biased when there is publication selection, and this is especially true for REE (Stanley, 2008; Moreno *et al*., 2009a; Stanley *et al*., 2010). Yet, the below "random-effects" PEESE-MRA estimates the "corrected" effects for both of these areas of research to be as large as or larger than REE and much larger than FEE. It seems clear that the reason for the large difference between the WLS and "random-effects" PEESE-MRA in Table 5.2 is that the random effects are picking up much of the publication selection bias. For both areas of research these differences are about threefold and have practical policy significance for any number of public projects and regulations.

Medical research does not have a culture of systematically explaining heterogeneity. Because most of their estimates come from controlled experiments, their priors are to assume that research variation is entirely random, and there are often too few comparable experiments to undertake a serious statistical study of heterogeneity. Nonetheless, there are differences in populations, strength of stimulus (dosage or treatment protocols), and experimental design that have systematic effects on the targeted health outcomes. Yet, medical researchers

Table 5.2 WLS and "random-effects" PEESE

Variables	*Statistical life*	*Price elasticity*
WLS of MRA (4.4)	1.665 (5.50)*	−0.115 (−7.76)
Random effects	5.70 (5.90)	−0.321 (−12.32)
n	39	110

* *t*-values are reported in parentheses.

tend to ignore these systematic effects and assume that any observed excess heterogeneity is random. This is the reason why STATA's **metareg** routine automatically includes random effects. However, in economics and business as well as in medical research, random-effects MRAs are likely to reintroduce publication biases that were carefully filtered by our simple WLS-MRA models of publication selection. Although this area of research merits further study, we believe that it is unwise to use random effects in meta-analysis.

5.2 Multivariate models of research

5.2.1 Meta-regression of publication selection

Publication selection causes incidental truncation from the population of estimates (Wooldridge, 2002). It is "incidental" because the estimates themselves are not directly selected, but rather their sign and statistical significance. The problem of incidental truncation and hence publication selection may be regarded as a special case of sample selection (Davidson and MacKinnon, 2004: 486–9). The conventional solution to this selection bias is to employ a "Heckman" two-equation system:

$$\mathbf{e} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5.2}$$

$$P = 1[\mathbf{K}\boldsymbol{\delta} + \mathbf{u} \geq 0] \tag{5.3}$$

where **e** is a vector of estimated effects, and $P_i = 1$ if e_i is reported in the literature and zero otherwise (Wooldridge, 2002: Chapter 17; 2006: 618–20). **Z** and **K** are matrices of exogenous variables. **K** affects the likelihood of selecting an empirical estimate, and **Z** models the heterogeneity and misspecification biases of the reported estimate. **ε** and **u** are assumed to be normally distributed with correlation ρ (Davidson and MacKinnon, 2004: 486–9). In typical economics applications, equation (5.3) is estimated by a probit using the entire sample of selected and non-selected observations. However, in the case of publication selection, we do not generally have access to unreported estimates.[6] Thus, step one of the conventional Heckman (1979) two-step method cannot be estimated for the publication selection of empirical economic research.[7] As a result, conventional Heckman selection corrections are not possible for application to publication selection.

Instead, incidental truncation gives an expected value quite similar to the conventional "Heckman" regression:

$$E(effect_i \mid truncation) = \alpha_1 + \sigma_i \cdot \lambda(c) \tag{5.4}$$

where $\lambda(c)$ is the inverse Mills ratio, $c = a - \alpha_1/\sigma_i$, α_1 is the "true" regression coefficient or empirical effect, a is the critical value of the standard normal distribution, and σ is the standard error (see Section 6.3 for a more detailed explanation and rigorous derivation of this and related relations). Replacing the inverse Mills ratio term in (5.4) with $\beta_1 SE_i$ gives our previously reported FAT-PET-MRA:

$$effect_i = \beta_0 + \beta_1 SE_i + \varepsilon_i \tag{4.1}$$

Recall that $\beta_1 SE_i$ represents systematic selection for statistical significance and provides a linear approximation of this truncation relation. The telltale signal of publication selection is a systematic relation of reported effects with their standard errors as revealed by meta-regression analysis (Card and Krueger, 1995a; Stanley, 2005a, 2008).

Both selection and authentic empirical effect are likely to be more complex than the simple models introduced in Chapter 4. Both terms in the simple FAT-PET-MRA (4.1) can be expanded to allow for greater complexity. The true effect, α_1, may be replaced by $\beta_0 + \sum \beta_k Z_{ki}$ to allow for heterogeneity and/or large-sample misspecification biases. Again, Chapter 6 discusses these issues in greater detail. Simple publication bias, $\beta_1 SE_i$, may also be given a multivariate form, $\beta_1 SE_i + \sum \delta_j SE_i K_{ji}$. Exploding MRA model (4.1) to allow both types of effects gives

$$effect_i = \beta_0 + \sum \beta_k Z_{ki} + \beta_1 SE_i + \sum \delta_j SE_i K_{ji} + \varepsilon_i \tag{5.5}$$

As a result, there will be no single "true effect", and publication selection will no longer be represented by a single term, but rather all the terms $\beta_1 SE_i + \sum \delta_j SE_i K_{ji}$.

$SE_i K_{ji}$ represents any factor that might affect the researchers' decision to report a given estimate. No doubt such factors will include the perceived quality of the econometrics used, such as whether the model includes obvious important variables (e.g. income in a demand relation), whether suspected econometric problems (e.g. non-stationarity) are properly accommodated, and whether appropriate econometric methods are employed. These K-variables may include any *observable* dimension of "quality." Recall our discussion of research quality in Chapter 2. The Z- and K-variables in MRA model (5.5) can be employed to accommodate the effects of research quality on both the magnitude of the actual empirical effects, Z, and the propensity to report an estimate, K.

Of course, as discussed in Chapter 4, MRA model (5.5) will also have heteroskedasticity. Either a WLS statistical routine should be used with precision squared, $1/SE_i^2$, as the "analytic" weights, or OLS can be applied on the MRA model that results from dividing equation (5.5) by the estimated standard errors, $SE_{i.}$,

$$t_i = \beta_1 + \sum \delta_j K_{ji} + \beta_0/SE_i + \sum \beta_k Z_{ki}/SE_i + u_i \tag{5.6}$$

where t_i is the reported t-value of the ith reported effect.

This multiple MRA model provides a flexible framework in which to explain the wide variation routinely found among reported research results. However, the structure of incidental truncation does not constrain (5.5) or (5.6) to be linear in SE, and simulations have show that using SE_i^2 provides a less biased, corrected estimate in a simple MRA; recall PEESE from Chapter 4. Thus, we will also investigate the alternative WLS-PEESE version of this multiple MRA model:[8]

$$t_i = \beta_1 SE_i + \sum \delta_j \mathrm{K}_{ji} SE_i + \beta_0/SE_i + \sum \beta_k Z_{ki}/SE_i + u_i \tag{5.7}$$

5.2.2 Multiple meta-regression of economics research

The above MRA models offer a broad framework to investigate the heterogeneity and selection of reported research results. Below, we discuss typical moderator variables found useful in past meta-analyses with particular focus on the employment effect of minimum wages and the value of a statistical life from hedonic wage equations.

Moderator variables

In Section 2.2, we discussed the typical dimensions of research that are routinely coded in meta-analyses. Chapter 2 identifies the standard error (precision) or the sample size as *essential* moderator variables to weight the reported estimates of effect and to correct for publication bias. *Typical* moderator variables include omitted relevant variables, alternative measures of key variables, econometric model and methods employed, and the data source used, among others. These typical moderator variables become “essential” if the MRA is to avoid potential misspecification biases due to omission of important explanatory variables.

Meta-regression analysis was originally conceived to quantify objectively the magnitude of likely misspecification biases in econometric applications (Stanley and Jarrell, 1989). Central among such biases is *omitted-variable bias*. Because economics and business research is typically observational, using pre-existing databases, researchers are often forced to exclude important explanatory variables from their models. Consequently, there is the real risk of misspecified studies. Fortunately, MRA can quantify the extent of likely misspecification bias.

For example, in a meta-analysis of the gender wage gap, whether the worker’s wage equation accounted for the workers’ age, experience, industry, and private/governmental job status were all found to affect the reported estimated wage gaps (Stanley and Jarrell, 1998). For minimum-wage studies, Doucouliagos and Stanley (2009) coded for whether the original model included workers’ education (*School*), the unemployment rate (*Un*), a time trend (*Time*), year-specific effects (*Yeareffect*), and regional effects (*Regioneffect*) – see Table 5.3 for the full list of variables.[9] Bellavance *et al*. (2009) find that whether compensation and endogenous risk are taken into account in the hedonic wage equation has important effects on the estimated value of a statistical life, while the omission of injuries does not seem to have a noticeable effect. See Table 5.4 for a list of moderator variables used in the meta-analysis of the VSL. Further discussion of these coded moderators is given below.

Alternative measures

Often the most important explanatory variables in a meta-analysis concern alternative ways that key variables are measured by the researcher. In the estimation of gender wage discrimination, the most important research dimension is how workers’

Table 5.3 Moderator variables for minimum-wage research

Moderator variable	*Definition*	*Mean (standard deviation)*
SE	standard error of the reported estimated elasticity	0.16 (0.39)
Panel	= 1 if estimate relates to panel data with time series as the base	0.45 (0.50)
Cross	= 1 if estimate relates to cross-sectional data	0.13 (0.34)
Adults	= 1 if estimate relates to young adults (20–24)	0.14 (0.35)
Male	= 1 if estimate relates to male employees	0.07 (0.26)
Non-white	= 1 if estimate relates to non-white employees	0.05 (0.22)
Region	= 1 if estimate relates to region-specific data	0.10 (0.30)
Lag	= 1 if estimate relates to a lagged minimum-wage effect	0.13 (0.34)
Hours	= 1 if the dependent variable is hours worked	0.07 (0.25)
Double	= 1 if estimate comes from a double-log specification	0.42 (0.49)
AveYear	the average year of the data used, with 2000 as the base year	−19.17 (11.90)
Agriculture	= 1 if estimates are for the agriculture industry	0.01 (0.11)
Retail	= 1 if estimates are for the retail industry	0.08 (0.27)
Food	= 1 if estimates are for the food industry	0.13 (0.34)
Time	= 1 if time trend is included	0.37 (0.48)
Yeareffect	= 1 if year-specific fixed effects are used	0.30 (0.46)
Regioneffect	= 1 if region/state fixed effects are used	0.34 (0.47)
Un	= 1 if a model includes unemployment	0.56 (0.50)
School	= 1 if model includes a schooling variable	0.15 (0.35)
Kaitz	= 1 if the Kaitz measure of the minimum wage is used	0.40 (0.49)
Dummy	= 1 if a dummy variable measure of the minimum wage is used	0.17 (0.38)
Published	= 1 if the estimate comes from a published study	0.85 (0.35)

Table 5.4 Moderator variables for hedonic estimates of the value of a statistical life

Moderator Variable	*Definition*	*Mean (standard deviation)*
SE	the standard error of VSL in millions of 2000 US dollars	3.02 (3.91)
AveIncome	the average income in thousands of 2000 US dollars	29.30 (9.50)
LnIncome	the logarithm of average income	10.20 (0.50)
Death	the average probability of death times 10,000	2.05 (2.52)
Year	year of publication, with 2000 as the base year	−9.56 (7.92)
EndoRisk	= 1 if the hedonic wage eq. uses an endogenous measure of risk	0.13 (0.34)
Comp	= 1 if the wage eq. includes compensation insurance	0.21 (0.41)
US	= 1 if the study used US data	0.54 (0.51)
UK	= 1 if the study used UK data	0.10 (0.31)
White	= 1 if VSL estimate relates to white workers	0.11 (0.31)
Union	= 1 if VSL estimate relates to union workers	0.16 (0.37)
SOA	= 1 if the data come from the Society of Actuaries	0.11 (0.31)

wages are measured: hourly wages, weekly earnings, annual salary or computed from annual salary. A third of the observed variation in reported gender wage gaps can be explained by how workers' wages are measured (Stanley and Jarrell, 1998). Likewise, how the minimum-wage variable is measured (using the Kaitz index, which accounts for changes in the effective minimum wage) is also found to be important – see below and Doucouliagos and Stanley (2009). Other measurement variables coded for the minimum-wage literature include hours worked (*Hours*) rather than employment, measuring the minimum-wage by a dummy variable (*dummy*), or using a lagged minimum-wage effect (*Lag*) – see Table 5.3.

Econometric model and methods

Often there are differences in the specification of the econometric model employed or in the econometric methods used to account for various dimensions of the causal structures or error terms. In economics research, there are usually differences in the functional form of the econometric model. Typically, meta-analysts code for whether or not a logarithmic specification is used (e.g. *Double* in Table 5.3). Models of panel data can be considered a different model specification (*Panel*, Table 5.3) or a difference in the type of data. Among estimates of the efficiency-wage effect on productivity, models that accounted for the likely simultaneity between wages and productivity were found to be much larger (Krassoi-Peach and Stanley, 2009). Studies that used Heckman selection correction methods in the gender wage literature were also found to be associated with much larger estimates (Stanley and Jarrell, 1998).

Data sources

Virtually all meta-analyses in economics have coded for the obvious potential heterogeneity caused by nuances in different sources of data. Surprisingly, different data sources do not always cause a noticeable difference to the reported estimate of effect, for example among estimated gender wage gaps (Stanley and Jarrell, 1998). However, differences in data and particularly subpopulation are found to be important in minimum-wage and VSL research, see below. Because it is conventional scientific practice, business and economic researchers are almost always careful to state and describe the source of their data. Thus, meta-analysts would be remiss if they did not code these data differences. One should be especially careful to denote the subpopulation, country or region, time period and industry from which the data are drawn, because these dimensions often contain unobservable, but important, differences in socio-economics or institutional settings not fully incorporated into the econometric model employed.

Value added

Unlike conventional narrative reviews or conventional applied econometric research, meta-analysis can add new and relevant information, unavailable to the original study. It is now conventional practice among meta-analysts to include several such moderator variables. Examples are legion. Take the value of a statistical life. Bellavance *et al.* (2009) recognize that the average income level of the workers studied (*AveIncome*, Table 5.4) may have an important effect on their willingness to accept higher risks of death. More affluent workers may be expected to hold out for greater compensation to increase their occupational risks. The average income of workers between different studies can be considerable because these studies involve different samples of workers from different jobs, countries, regions, or time periods. Such variation in *average* income cannot be controlled by the individual econometric study, because it does not vary within a study but only varies across studies. If a meta-analysis did nothing more for our understanding of the VSL than account objectively and systematically for this one important dimension, it would make an important contribution. Other study-invariant factors coded by Bellavance *et al*. (2009) are the year a study was published (*Year*) and the average probability of death (*Death*). For the minimum-wage effect on employment, the average year of the data (*Aveyear)* is coded because it has been alleged that there have been structural changes to the relationship between minimum wage and employment (Doucouliagos and Stanley, 2009).

Because economic and business research is a socio-economic enterprise, meta-regression has the potential to account for these factors as well. For example, Stanley and Jarrell (1998) found that the gender of the researcher is correlated with the reported estimate of the gender wage gap.[10] Funding source and professional associations have also been found to be important (Doucouliagos and Laroche, 2003; Doucouliagos and Paldam, 2008). Regardless of the judgment of meta-analysts, meta-analysis is empirical; thus, the research record itself will decide which research dimensions are relevant. As we have discussed earlier (Chapter 2), it is important to err on the side of inclusion and code any research dimension that is suspected to have an important effect of the reported results.

5.3 Illustrations of multiple meta-regression analysis

There have been hundreds of multivariate meta-analyses conducted in economics and business.[11] We have chosen to focus on only two of the previously discussed illustrations, the VSL and the employment effect of minimum-wage increases, due to considerations of space and the richness of the moderator variables that have been coded. The discussion of the MRA of the VSL literature in Section 5.3.1 draws heavily from Doucouliagos *et al.* (2012b), and the investigation of the minimum-wage literature in Section 5.3.2 draws heavily from Doucouliagos and Stanley (2009).

5.3.1 The value of a statistical life

Bellavance *et al*. (2009) investigate the sources of the wide variation observed among the reported estimates of the value of a statistical life from hedonic wage equations. Among the explanatory variables that they code are the endogenous nature of risk, the presence of compensation insurance, the average income of the sample used, and the population from which the sample is taken. However, they did not test or control for publication selection. Accommodating and correcting for potential publication bias is the central contribution of Doucouliagos *et al*. (2012b) to this important literature. As we have shown in Chapter 4, there is strong evidence of publication bias among these reported estimates of VSL. The estimated β_1 from the simple FAT-PET-MRA (Table 4.1) is quite large, 3.2, easily allowing us to reject the null hypothesis of no publication selection (reject H_0: $\beta_1 = 0$; $t = 6.67$; $p < 0.001$). Clear publication selection is also seen in the associated funnel graph (see Figure 4.4 or Figure 5.1).[12] Thus, publication selection must also be included in any explanatory multiple MRA. Otherwise, the explanatory MRA would itself be subject to omitted-variable bias. Because this omitted-variable bias places Bellavance *et al*.'s (2009) findings in question, we conduct our own multiple meta-regression analysis.

Table 5.4 lists the coded moderator variables that might potentially explain the large variation in reported VSL. If each variable is allowed to be either a *Z*- or *K*-variable, we have 23 variables to explain 39 VSL estimates. For multiple MRA, we recommend using a general-to-specific (G-to-S) approach, or backward selection. There are always many research dimensions that might potentially affect the magnitude of the reported results. Thus, the number of possible MRA models routinely exceeds the number of observations, and the inflation of type I errors virtually guarantees that several research dimensions will be seen as statistically significant even if the results were only random noise (Sala-i-Martin, 1997).

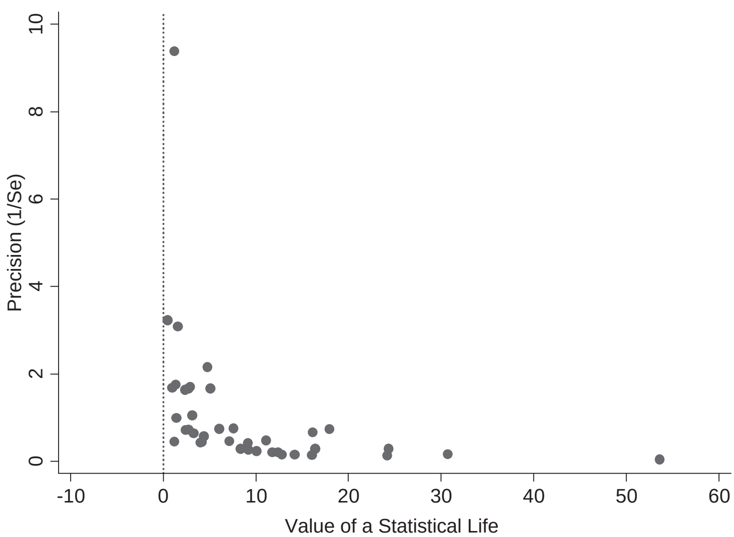


Figure 5.1 The value of a statistical life (in millions of 2000 US dollars)

Source: Bellavance *et al*. (2009).

When a mere 10 moderator variables are coded and are allowed to be either *Z*- or *K*-variables, there will be more than a million possible models. Only coding 10 moderators represent a rather modest meta-analysis in economics.

The G-to-S approach begins with all explanatory variables included in the equation. Then the least statistically significant variable is removed, one at time, until only statistically significant variables remain. Yes, this too is less than ideal, but some choice of moderator variables is necessary. "The strength of general to specific modeling is that model construction proceeds from a very general model in a more structured, ordered (and statistically valid) fashion, and in this way avoids the worst of data mining" (Charemza and Deadman, 1997: 78). The only other sensible approach is to report only the MRA model that includes all coded moderator variables. Any other model has to be regarded as having "negative" degrees of freedom (hence essentially worthless), because the number of meta-regressions considered and the selection mechanism could be anything. Unfortunately, this general, all-inclusive, MRA model also has great limitations. Assuming that there are sufficient degrees of freedom to run the all-inclusive MRA, the fog of high multicollinearity and low statistical power virtually guarantees the obscuring of much of the existing pattern of research.

Returning to the VSL example, three variables (*LnIncome*, *Year*, and *Death*) that are constant for each study are selected to be only *Z*-variables (i.e. research dimensions that explain the variation among reported estimates) because it is unlikely that researchers were selecting across these dimensions to obtain significantly positive VSL estimates. Meta-analysts may restrict some of the MRA coefficients in equation (5.6) to be zero when they have reason to do so. Here, we have so few observations of VSL available (39) that we believe that it is prudent to reduce the number of *Z*-/*K*-variables. Adding dozens of interaction terms (*K*-variables) will almost certainly create high multicollinearity and thus obscure individual conditional effects. We suggest that meta-analysts constrain some of the δ_j coefficients in MRA (5.6) to be zero whenever there is a "theoretical" reason to do so.

The remaining moderator variables are entered as both *Z*- and *K*-variables (i.e. those that allow for a differential propensity to report a given VSL estimate). Beginning from a multiple MRA that contained 20 moderator variables and the WLS-MRA version of equation (5.5), the G-to-S approach identifies four variables as statistically significant (see Table 5.5). Note that Table 5.5 is reported in the form of MRA model (5.5) and is estimated by using a WLS statistical package with $1/SE_i^2$ as the analytic weights. The OLS version of MRA model (5.6) will give the same results.[13]

Although many variables were allowed to be *K*-variables, G-to-S modeling identifies *SE* alone to be related to publication selection. The MRA coefficient for *SE*, 3.07, is in millions of US dollars and suggests that if the standard error of the VSL estimate increases by $1 million, reported VSL will increase by $3.07 million, *ceteris paribus*. This is a huge effect, practically and statistically, inflating the average estimate of VSL by $9.15 million and explaining nearly half of the observed variation among reported VSL estimates.[14] Selection dominates the reported hedonic wage estimates for the value of a statistical life.

Table 5.5 General-to-specific multiple MRA of the value of a statistical life (dependent variable = VSL in millions of 2000 US dollars)

Moderator variables	*WLS-MRA (5.6)*	*Robust MRA (5.6)*	*PEESE*
Intercept	−15.8 (−2.11)*	−31.6 (−2.58)*	−31.5 (−3.99)*
LnIncome	1.86 (2.28)	3.36 (2.70)	3.63 (4.20)
Year	0.19 (3.36)	0.18 (3.76)	0.21 (3.18)
Comp	−1.88 (−2.20)	−1.52 (−2.45)	−2.71 (−2.71)
SE	3.07 (5.12)	2.80 (6.55)	–
SE²	–	–	0.28 (2.78)
n	39	39	39
Adj R²	0.58	–	0.40
Standard Error	1.3	–	1.6

* *t*-values are in parentheses.

Source: Doucouliagos *et al*. (2012b).

Table 5.5 finds three *Z*-variables (*LnIncome*, *Year* and *Comp*) related to the observed heterogeneity among these reported VSLs, and their coefficients are quite reasonable. We would expect that life is a normal good and that workers value their lives more highly as their income increases. This is confirmed by our meta-regression ($p < 0.01$). The positive coefficient on *LnIncome* corroborates this expectation, and its value implies that if average income were to increase by 1 in natural logs, or 2.72 times, workers behave as if their lives are worth \$1.86 million more. We also find a trend among these estimates that increases VSL by \$190,000 per year. Note that both of these variables (*Year* and *LnIncome*) constitute "value added" by the meta-regression. That is, these moderator variables are study-invariant; thus, their influence on VSL cannot usually be investigated by conventional econometric analysis.[15]

The MRA coefficient on *Comp* is also quite sensible. Studies that control for worker's compensation insurance (*Comp* = 1) tend to report \$1.88 million lower VSL estimates. Thus, we corroborate the common understanding in this research literature that the failure to account for workers' compensation insurance can cause a significant exaggeration to the VSL.[16]

Lastly, the estimated intercept may seem somewhat absurd. How can the value of a life be nearly *negative* \$16 million? This meaningless number represents an extrapolation of the MRA estimated equation to an average income of \$1 (recall that the natural log of 1 is 0), well outside the observed range of average income (\$3,000 to \$49,000). Needless to say, intercepts of economic relations are often meaningless or misleading because they refer to very implausible and irrelevant circumstances. If the intercept were recalibrated to the average observed *LnIncome* (10.2), it would become 4.3, implying that VSL would be \$3.3 million at the average log income for the year 2000 when the hedonic wage equation did not account for worker compensation.

Because economics and business research is conditional, determining a representative estimate of effect requires some judgment on the part of the meta-analyst and some notion about what constitutes "best practice" research (Doucouliagos and Stanley, 2009). Recall that when we ignore the multidimensional nature of VSL research, our corrected estimate, of VSL is $1.67 million with a 95 percent confidence interval of (1.05, 2.27) – see PEESE Table 4.2. However, this simple MRA does not allow for the effects of the other moderator variables found important above. But what values should we substitute into this multiple MRA for these moderator variables? In conventional econometric applications, researchers will often use the sample means of the independent variables to avoid making such judgments. However, this makes no sense in the context of meta-analysis. Using the sample means for all of our moderator variables will just give us back the average reported VSL. But we already know that this average contains much publication bias and perhaps misspecification biases as well. Surely meta-analysis can do better.

First, we need to remove identified publication selection bias by setting $SE = 0$. Recall that, as $SE \rightarrow 0$, a study approaches perfection with no estimation error and no publication bias. Selecting a year is easy and simply depends on the year for which we wish to estimate or to use as an arbitrary benchmark. We choose 2000 because it is a nice round number. Less obvious professional judgment is required to choose the appropriate values of *Comp* and *LnIncome*. As discussed above, it is widely argued in this research literature that omitting worker compensation biases results upward. Thus, following best practice in this literature, we substitute 1.0 for *Comp*. Alternatively, using the sample mean, 0.21, would bias the estimated VSL upward by about $395,000. Lastly, what is the most appropriate value of worker income? The answer depends on the specific group of workers that one wishes to use as a reference group. For our current purposes, we use the sample average worker log income, because we seek only to provide a generally representative VSL estimate corrected for identified biases.

Next, we substitute these values into the WLS multiple MRA model reported in column 1 of Table 5.5. Doing so "predicts" VSL to be $1.36 million, with 95 percent confidence interval ($34,000, $2,693,000).[17] This prediction is quite close to the simple PEESE estimate that did not explicitly control for these other dimensions of research. Even the downwardly biased PET coefficient (Table 4.1) is well within this confidence interval, and vice versa. In Section 6.4.1 we argue that the simple MRA models of publication selection introduced in Chapter 4 may reflect the total publication bias despite potential omitted-variable bias, and their estimates often adequately summarize a research literature. This interpretation is consistent with the multiple MRA of the value of a statistical life.

It is customary in applied econometrics to run a few "robustness checks," and meta-analysts will likely be forced to follow this practice if they expect their research to be published in top economic journals. Especially worrisome is that one study (Sandy and Elliott, 1996) reports a VSL more than twice that of any other, approximately $54 million. Because this study also has the largest standard error, these MRA models will give it a small weight and a larger publication bias correction – recall our discussion in Chapter 4. Nonetheless, it would be prudent to use robust regression methods that minimize the influence of any one potentially influential outlier. Column 2 of Table 5.5 reports the robust regression version of our WLS-MRA and give virtually the same results, except that the effect of higher income is notably larger.[18] The MRA robust regression coefficients predict VSL to be $1.15 million in 2000 for a worker with the sample of average income when worker compensation has been included.

As a further check, we also report the multiple PEESE-MRA (recall model (5.7)) in column 3 of Table 5.5. Using the variance, $\beta_1 SE^2$, to approximate publication selection bias gives very similar statistical results. However, a corrected VSL based on this PEESE-MRA is somewhat higher, $2.74 million ($1.30 million, $4.18 million), but like our other estimates is much smaller than the simple average reported in this research literature ($9.48 million).

When there are multiple estimates reported per study, other MRA models are needed to control for the potential dependence with studies and to ensure the robustness of simpler MRA structures. Because Bellavance *et al*. (2009) did not collect multiple estimates from each study, there is no need to accommodate potential dependence among these VSL estimates. However, the minimum-wage research literature routinely reports many estimates per study, and we discuss the use of appropriate multiple MRA models of within-study dependence in the next section. Before we move on to the minimum-wage literature, we should note that there has been much consistency among the simple MRA models of publication bias for VSL and their multiple MRA counterparts. The central findings are that VSL is greatly reduced when identified publication selection is filtered from the reported estimates, rises with income, and is reduced further by nearly $2 million when worker compensation is included. See Doucouliagos *et al*. (2012b) for a further discussion of this meta-analysis of VSL estimates and its policy implications.

5.3.2 The employment effect of raising the minimum wage

Doucouliagos and Stanley (2009) extend Card and Krueger's (1995a) controversial meta-analysis of the minimum wage and corroborate their central findings: the minimum wage has no genuine adverse employment effect and there is much selective reporting of an adverse effect. Recall the minimum wage's asymmetric funnel graph (Figure 4.6, repeated as Figure 5.2, below) and the clear MRA evidence of publication bias (reject FAT, H_0: $\beta_1 = 0$; $t = -4.49$; $p < 0.001$ – see Table 4.1). Yet, the most important finding of this extensive meta-analysis of the minimum wage's employment effects is that after proper allowance for publication selection is made no adverse employment effect remains.

As discussed earlier, Table 5.3 lists the 22 coded *Z*-/*K*-variables that can be used to explain the large variation of minimum-wage effects. This list of MRA moderator variables was determined purely by the type of data available and by debates in the minimum-wage literature. Because some estimates are industry-specific, we control for possible industry differences in employment effects (*Agriculture*,

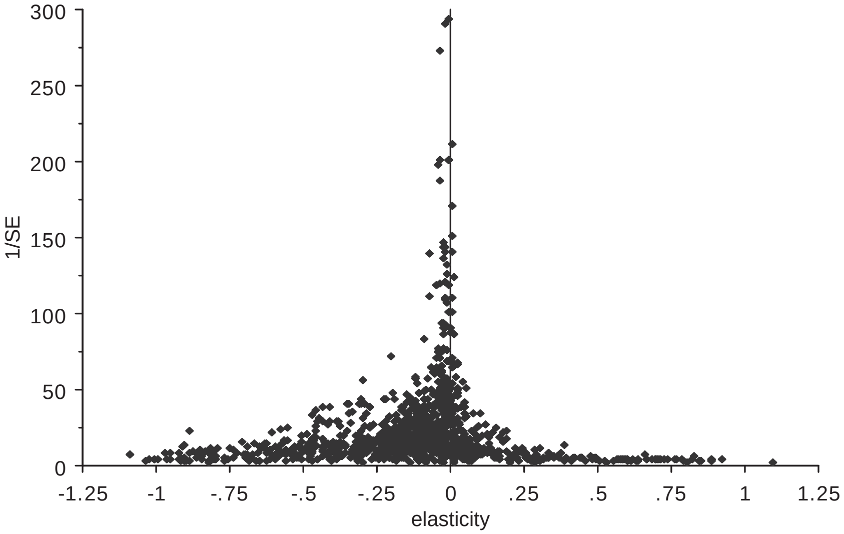


Figure 5.2 Funnel graph of estimated minimum-wage effects (n = 1,424)

Source: Doucouliagos and Stanley (2009).

Retail, and *Food*) for estimates relating to agriculture, retail, and food (mainly restaurants). Typical subpopulation concerned the differences between males and females (*Male*), whites and non-whites (*Non-White*), and teenagers and young adults (*Adults*). Some estimates relate to a specific region, and the variable *Region* is included to control for any differences between region-specific and US-wide elasticities. Although most estimates relate to the contemporaneous employment effects of a minimum-wage rise, many estimate the lagged effect of minimum wage rises (*Lag*).

There is some debate in the literature about the need to control for cyclical effects (*Un*) and school enrollment (*School*). These are included to investigate whether omitting them creates any noticeable “bias” or differences. The vast majority of estimates relate to employment, but some relate to hours worked (*Hours*); such differences in the measure of employment might affect the estimated elasticity. *Time* allows for any effect of including a time trend in the specification of the employment equation. A large group of estimates (696) come from studies that use panel data (*Panel*), while 210 use cross-sectional data (*Cross*). Two related variables are *Yeareffect* and *Regioneffect*, which control for the inclusion of period and cross-section (region/state) fixed effects, respectively.

The majority of minimum-wage elasticity estimates have been reported in published academic journals (*Published*), while others come from working papers and have yet to be published. Two final controls relate to further differences in the measurement of the minimum wage – the use of the Kaitz index (*Kaitz*) and the use of a dummy variable (*Dummy*) for the presence of minimum wage.

As before, we used a G-to-S approach, and the resulting multiple MRA is reported in Table 5.6. With 1,474 estimates of the employment elasticity of the minimum wage, there are ample degrees of freedom for this G-to-S modeling approach; thus, we allowed all of these moderator variables to be both *Z*- and *K*-variables. Furthermore, we did not have *a priori* reasons to constrain any of the possible effects to be zero. The G-to-S process resulted in 14 variables remaining statistically significant, 12 *Z*-variables and two *K*-variables, and explaining 41 percent of the reported variation among minimum-wage elasticities.[19]

Of special note are the significant time trend, *Aveyear*, which suggests that adverse employment effects, if any, are smaller each year, and the large MRA coefficient on *Panel*. The latter suggests that studies using panel data find employment elasticities to be approximately 0.18 more negative, or adverse, than those using time series data. However, this must be seen in the full context of the multiple MRA reported in Table 5.6. Note that the intercept, 0.12, suggests that a 10 percent increase in the minimum wage actually *increases* employment by 1.2 percent, assuming of course all the other MRA moderator variables are zero. When a study uses panel data and nothing else is coded "1", the WLS-MRA estimates the minimum-wage elasticity to be −0.062, which is much less than the average reported elasticity, −0.19. But this assumes that all of the other moderator variables are zero, and this is not consistent with a high-quality research study. To have only *Panel* = 1 implies that study did not include any year effects (0.069), did not use the Katz index (0.052) or a log-log specification (0.064), among many other things. One could easily argue that the better studies include year effect, uses the Katz index and a double-log specification. When included with a panel study, these research dimensions predict a positive employment effect from minimum wage raises (0.123).[20]

Publication selection

For the minimum-wage literature, in a multiple MRA context, neither publication bias nor authentic effect is represented by any single MRA coefficient. These effects are themselves multivariate. In particular, the MRA coefficient on *SE* is no longer a measure of the magnitude of the average publication bias by itself. Rather, it is the combination of this MRA coefficient and all the *K-variables* (*Un·SE* and *Double·SE*). We can easily test the joint hypothesis that all of the associated MRA coefficients are zero. Doing so gives clear evidence that publication selection remains in this multiple MRA ($F_{(3,\ 1459)} = 84.2$; $p < 0.0001$).[21]

Estimated MRA coefficients from these *K*-variables can be used to calculate the average estimated publication bias for a given research literature, which is −0.218 for the minimum-wage literature (calculated in terms of employment elasticity). This is quite comparable to the value, −0.256, obtained from the simple MRA (Table 4.1).[22] Subtracting either of these estimated publication biases from the reported minimum-wage elasticities converts the average minimum-wage elasticity (−0.190) to a small *positive* value. Regardless of its sign, this positive value is so small that it is of no practical import.

Table 5.6 Multiple MRA of minimum-wage research: WLS of model (5.5)

Moderator variables	*Column 1: WLS*	*Column 2: Cluster-robust*	*Column 3: REML*	*Column 4: FEML*	*Column 5: Robust*
Heterogeneity (Z-variables)					
Intercept (β_1)	0.12 (10.45)*	0.12 (4.39)*	0.11 (7.03)*	0.10 (6.04)*	0.084 (8.78)*
Panel	−0.18(−18.60)	−0.18 (−4.72)	−0.15 (−12.38)	−0.15 (−10.48)	−0.19 (−23.59)
Double	0.064 (8.90)	0.064 (3.20)	0.044 (5.98)	0.041 (5.42)	0.033 (5.45)
Region	0.040 (3.28)	0.040 (0.92)	0.087 (6.36)	0.090 (6.34)	−0.065 (−6.43)
Adult	0.024 (4.27)	0.024 (2.68)	0.021 (3.75)	0.021 (3.72)	0.019 (4.04)
Lag	0.026 (4.46)	0.026 (1.60)	0.012 (2.08)	0.010 (1.59)	0.010 (2.13)
AveYear	0.004 (11.86)	0.004 (4.34)	0.003 (7.44)	0.003 (6.38)	0.003 (9.20)
Un	−0.042(−6.47)	−0.042 (−3.04)	−0.041 (−6.15)	−0.042 (−5.79)	−0.020 (−3.82)
Kaitz	0.052 (8.76)	0.052 (3.06)	0.034 (4.51)	0.032 (3.88)	0.025 (5.05)
Yeareffect	0.069 (8.61)	0.069 (1.98)	0.068 (7.84)	0.067 (7.44)	0.106 (15.80)
Published	−0.041(−7.85)	−0.041 (−2.69)	−0.039 (−5.63)	−0.037 (−4.89)	−0.028 (−6.48)
Time	−0.022(−3.95)	−0.022 (−2.08)	−0.020 (−3.10)	−0.017 (−2.46)	−0.013 (−2.85)
Publication selection (K-variables)					
SE	−0.36 (−0.26)	−0.36 (−0.11)	−1.21(−3.86)	−1.37 (−5.94)	0.11 (0.96)
Double·SE	−1.48 (−8.52)	−1.48 (−3.23)	−1.09 (−4.33)	−1.07 (−3.90)	−1.01 (−6.97)
Un·SE	−0.84 (−4.53)	−0.84 (−1.87)	0.84 (2.61)	1.16 (3.08)	−0.98 (−6.36)
No. of obs. (n)	1,474	1,474	1,474	1,474	1,474
No. of studies (k)	64	64	64	64	64

*t-values are reported in parentheses.

Source: Doucouliagos and Stanley (2009).

Heterogeneity of minimum-wage elasticities

Like publication bias, genuine heterogeneity is multivariate. Rather than any single overall minimum-wage effect on employment, there are many. The moderator variables *Panel*, *Double*, *Region*, *Adults*, *Lag*, *AveYear*, *Un*, *Kaitz*, *Yeareffect*, *Published* and *Time* all have a noticeable impact upon reported minimum-wage effects beyond publication selection. Testing whether all of these MRA coefficient are jointly zero, along with the intercept, β_1, finds easy rejection ($F_{(11, 1459)} = 50.8$, $p < 0.0001$). There are genuine, systematic patterns among reported minimum-wage research findings.

Box 5.2 Descriptive vs. explanatory MRA

There are alternative interpretations of what meta-regression does. One view considers the reported results to be the *population* of research in a given area of inquiry and seeks merely to describe this research population. By this view, it is enough to record the fact that research findings vary due to specific choices of methods, models, variables and data; reporting a descriptive summary of research and the associated response surface discharges the meta-analyst's scientific obligations. The second view is that reported research results are a sample from a virtually infinite population of possible findings that might be produced for a given phenomenon. From this second perspective, the purpose of MRA is to make inferences to the population of possible research results and to estimate what the conditional population mean would be under given research specifications. The obligation of the meta-analyst is then to explain the systematic variation observed among the reported findings and thus to identify the underlying response surface. But with inference as the objective, the obligations of the meta-analyst go further—to estimate the associated empirical effect for what might be regarded as best scientific practice. We take this second view, because we know that the reported findings in economics are a selected sample from a much larger set of produced results (Sala-i-Martin, 1997).

Estimating the corrected effect from a multiple MRA

So which of these significant effects represents the "true" employment effect of minimum wage? No single effect may be regarded as the authentic one, but by exercising some professional judgment the meta-analyst can determine the message of "best practice" research. As discussed previously, first we must filter out the publication selection, which implies that *SE* is zero. This makes the effect of all of the *K*-variables zero. Next, we need to substitute specific values for the *Z*-variables into the estimated MRA. To minimize any effect from potentially questionable judgment on our part, we first substitute the sample means of the *Z*-variables and use the current year (2012) for the average year of the data.[23] Doing so gives a corrected estimate of the employment elasticity of the minimum wage of +0.10 with 95 percent confidence interval (0.08, 0.12). Ironically, the consensus in the field is that the minimum-wage employment elasticity is about the same magnitude, but negative (−0.1). When proper allowance for publication bias is made, a small adverse employment effect becomes a small *positive* effect on employment!

To some, a positive employment effect might seem impossible because it flies in the face of a downward-sloping demand for labor that is taught in every introductory economics class. However, such a positive employment effect is an implication of efficiency wage theory (Akerlof, 1982; Card and Krueger, 1995b; Stanley and Doucouliagos, 2007). Furthermore, a separate meta-analysis of the empirical literature on efficiency wages strongly corroborates its existence (Stanley and Doucouliagos, 2007; Krassoi-Peach and Stanley, 2009). The findings from one meta-analysis may point to the need for further research to explain any unconventional or surprising finding by conducting another meta-analysis. Such learning by refutation and corroboration is how science is often described to progress (Popper, 1959, 1963). Nonetheless, we do not wish to claim that there is a genuine *positive* employment effect from raising the minimum wage. It is sufficient to find a clear and robust absence of an adverse employment effect to reject the applicability of the conventional theory of competitive labor markets to the US labor market.

A critic might correctly point out that using the sample mean values of the reported research base does not represent best practice in economics research. Fair enough, but what might represent the "best practice" research in this area? A case could be made that a published paper that uses panel data (including year fixed effects) and the Kaitz index is at least a part of "best practice."[24] When doing so, our MRA model (columns 1 and 2 of Table 5.6) predicts a reduction of the above positive employment effect to +0.060, with confidence interval (0.040, 0.081). What constitutes "best practice" in minimum-wage research is, however, somewhat debatable. For example, Burkauser *et al*. (2000) argue against the use of including fixed year effects in minimum wage studies. If these effects are removed from the best-practice calculation, our MRA model predicts a practically and statistically insignificant effect of −0.012 (−0.039, +0.015). On the other hand, Card and Krueger (1995b) argue for the inclusion of fixed effects but against the use of the Kaitz index. Adapting this definition of "best practice" to our MRA model predicts a very small positive, but insignificant, employment effect of +0.008 (−0.007, +0.024). Regardless of one's view of "best practice," no practically significant, adverse employment effect remains for the US labor market after correcting for publication selection bias.[25]

5.4 Robustness and dependence

Even if the weighted least-squares MRA specification were entirely correct, reviewers and editors will demand that its central findings be robust to reasonable model variations. Thus, robustness checks must always be conducted. Furthermore, as we discussed in Chapter 4, there is potential dependence among estimates of the same study when multiple estimates are reported by a given study. This potential dependence must be explored to ensure the WLS-MRA findings are valid.

We recommend two general modeling strategies to accommodate dependence among estimates and to correct the MRA's standard errors accordingly: cluster-robust and multilevel (or equivalently, unbalanced panels).[26] Reported estimates can be clustered by any dimension within which reported estimates are thought to be correlated. The dataset used by a study, the author of the study, and the study itself are reasonable dimensions upon which to cluster. After deciding on the clustering variable, statistical packages will calculate the cluster-robust standard errors from a generalized least-squares approach. Note that a cluster-robust MRA should give exactly the same MRA coefficients as the simple WLS-MRA (see Table 5.6). The only difference is that the standard errors are computed in a manner to account for any potential dependence among the estimates within the specified clusters.

Of the four example meta-datasets we have used thus far, only the minimum-wage literature has the full multidimensional data structure needed for panel or cluster analysis. Recall that we found 64 empirical studies that estimate the US minimum-wage employment elasticity, which jointly report 1,474 estimates along with their standard errors. Thus, the typical study in this literature engages in much robustness checking and reports 23 estimates of the minimum-wage elasticity. For our MRA of minimum-wage effects, we clustered by study. Clustering has little practical effect on the MRA of minimum-wage effects (see column 2 of Table 5.6). Of course, all of the MRA coefficients have different and generally smaller *t*-values, but the same research dimensions remain statistically significant with two exceptions, *Region* and *Lag*. Yet, even this slight difference has no effect on our central findings. We still find there to be important publication bias and systematic heterogeneity, and the restrictions tests still confirm this assessment. The only notable difference is that studies that use regional data and those that report a lagged employment effect may not be so different from the rest of this research literature. However, our "takeaway points" are not concerned with any specific factor that might affect reported minimum-wage effects, except perhaps *SE*. Rather, we wish merely to be sure that potential systematic heterogeneity is accounted for and thereby not allowed to bias our overall findings.

Multilevel modeling is equivalent to an unbalanced panel model, which is quite familiar to econometricians, in general, and especially those economists who estimate minimum-wage effects. Recall that 45 percent of the minimum-wage estimates come from panel models (Table 5.3). Rosenberger and Loomis (2000b) were the first to recommend the use of unbalanced panel methods to account for within-study dependence in the context of MRA, and this method has been advocated by many others (e.g. Bateman and Jones 2003). In the minimum-wage meta-data, the Durbin–Watson statistic, 0.94, reflects the presence of within-study dependence.

In Chapter 4, we discussed the structure and motivation for using multilevel models. Here, we include unobserved study effects in our *Z*/*K* multiple MRA (equation (5.6)),

$$t_{is} = \beta_1 + \sum \delta_j K_{jis} + \beta_0/SE_{is} + \sum \beta_k Z_{kis}/SE_{kis} + v_s + u_{is} \tag{5.9}$$

for the *i*th estimate in the *s*th study. v_s represents an unobserved study effect and can alternatively be replaced by a "fixed-effects" term, δD (where D is a matrix of study dummy variables). Equation (5.9) may also be regarded as a generalization of the multilevel MRA model discussed in Chapter 4 (equation (4.5)), but one that allows for heterogeneity and a more complex structure of publication selection. An alternative approach is to model unobserved study effects explicitly using WLS in the context of MRA (5.5); recall Chapter 4.

Table 5.6 reports both the random-effects multilevel (REML) MRA (column 3) and the fixed-effects multilevel (FEML) MRA (column 4), for minimum-wage research. REML is more accurately described as a mixed-effects model because it contains both the fixed effects of all *Z*- and *K*-variables along with random normal unobserved study effects. Although Table 5.6 is reported in the form of MRA model (5.5), the WLS-MRA, equation (5.6) or (5.9), is used in all cases. Here, too, the findings are largely the same as the previously discussed multiple WLS-MRA findings. Again, the overall results of large publication bias and of considerable systematic heterogeneity remain; however, their specific structure changes somewhat. All of the *Z*-variables, which map the structure of heterogeneity, remain statistically significant with the same signs, though their coefficients change a little. The only notable difference is that the *K*-variable, *Un·SE*, changes signs but remains statistically significant. Although we are not particularly interested in any individual effect of a moderator variable, this change poses a curious puzzle worth investigating further.[27]

The WLS-MRA finds that those minimum-wage studies that add the unemployment rate to the employment equation are more likely to select for adverse employment effects, *ceteris paribus*. This tendency is reversed when study-level effects are allowed. The explanation for this reversal comes from the existence of large outliers reported by a few studies. One study contains a dozen elasticities between nearly −5 and −10; a second study reports four elasticities approximately +2 and larger. Such large elasticities, whether positive or negative, are simply not plausible. When unobserved study effects are allowed, these outliers dominate the estimation of the study effects. Recall that the average reported elasticity is −0.19. One of the beauties of the WLS-FAT-PET-MRA is that estimates with large standard errors are given little weight, while precise estimates are given a much larger weight. All of the studies with implausibly large elasticities also have corresponding large standard errors; thus, the WLS-MRA automatically discounts them. FAT-PET-PEESE-MRAs are remarkably resilient to such outliers.

Nonetheless, to ensure our findings are robust to the influence of a few outliers, column 5 of Table 5.6 reports the associated robust multiple MRA. Robust regressions are designed to be resilient to outliers and leverage points. Note that robust regression MRA finds a significantly negative *Un·SE* coefficient, consistent with the WLS-MRA. But most importantly, the overall findings of significant publication bias and heterogeneity are confirmed.

5.5 Will the real meta-regression analysis model please stand up?

Thus far, we have been very optimistic about the prospects of employing meta-analysis to identify the major patterns in economic and business research. We have seen how meta-regression analysis can identify and correct for publication selection bias, and identify and estimate genuine empirical effects and misspecification biases in both theory and practice. However, we are less sanguine that MRA can accurately identify the full complex structure of economics or business research. Because there is an indefinite number of potential MRA models, even mature areas of research will have insufficient degrees of freedom to investigate them all fully. Furthermore, the research literature is likely to contain idiosyncratic research choices that influence the reported results. If some of these choices go unreported and are coincidentally correlated with moderator variables, a given moderator's MRA coefficient may represent some other unobserved research dimension. Thus, it is unlikely that the values of each reported empirical estimate can be fully understood by estimated MRA coefficients. Needless to say, even if we estimate the "true" MRA model, there would remain random unexplained research variation. But these limitations are not unique to meta-analysis; they also apply at least as fully to applied econometrics. Business and economic research is limited by the observational data available, which are likely to be influenced by unobserved, but important, variables, and the number of potential economic models is routinely larger than the number of observations (Sala-i-Martin, 1997).

Nonetheless, we remain confident that meta-regression analysis can adequately identify a few of the important characteristics of a research literature whether or not there is a practically significant effect. Recall that in the four meta-analysis examples used, two (the value of a statistical life and the price elasticity of the demand for water) found a genuine empirical effect in spite of strong publication selection bias. With union-productivity correlations, there is little sign of publication selection or genuine effect, and we find no adverse employment effect from the minimum wage after allowing for publication selection. All of these general findings are robust and confirmed by multiple MRA. Furthermore, much of the shape of research is revealed by MRA. In the hedonic wage estimation of VSL, average income, the presence of worker compensation for injuries and a time trend were robustly identified as important. All of these factors are consistent with theory. For minimum-wage research, many factors such as: the use of panel data, year effects, the Kaitz index, a double-log model, data from young adults, lagged effects and whether the study was published all had robust and consistent effects on the reported minimum-wage elasticity. Obtaining such a clear and objective understanding of any area of research is an important achievement.

But then, what is the best MRA model of economic research and how can we decide? In our view this is the wrong question to ask. The right question is: What research dimensions are robust to MRA model specification? However, econometric training will cause many researchers to search for the "best" model. Fortunately, there are objective criteria for MRA selection.

To illustrate the issues at stake, we return to our richest example of meta-analysis, minimum-wage employment effects. Recall that this example has a multilevel structure where many estimates are typically reported by each of the 64 studies. Due to concerns of dependencies among estimates within a study, best econometric practice would suggest that we use either cluster-robust or a panel model. But which one?

First, let us consider whether there are significant study-level effects. For this purpose, one can use the Breusch–Pagan Lagrange multiplier (LM) test. STATA calculates this to be 1,272 for the minimum-wage literature, which is significant at any level.[28] Clearly, there are study-level effects. But are they "fixed" (FEML) or "random" (REML)?

To answer this second question, the generic Hausman specification test can be used (Hausman, 1978). As with all applications of the Hausman test, it is able to differentiate between an estimator that is consistent under both alternative specifications (FEML for this application) and a second estimator that is consistent but also efficient only under one specification (REML). Recall that all "random-effects" models assume that the random effects are uncorrelated with the independent variables. However, as we discussed previously in this chapter, publication selection is likely to make selected random effects correlated with the standard error, which is one explanatory variable needed to be investigated in any MRA application. This correlation is easily seen in Monte Carlo simulations and can, in effect, be tested by the Hausman test. Returning to the minimum-wage example, the Hausman test gives $\chi^2_{(13)} = 36.62$ ($p < 0.001$) and clearly rejects the random-effects MRA in favor of the fixed-effects version. As expected, a random-effects multilevel MRA is misspecified. The interested reader should consult Feld and Heckemeyer (2011), especially Figure 2, for a more comprehensive diagram and discussion of model specification testing for MRA. Feld and Heckemeyer (2011) also give an excellent comprehensive discussion of the econometric issues at stake in MRA modeling.

What might surprise some applied researchers, especially medical researchers, is that the fixed-effects model is a generalization of the random-effects models because it allows the study-level effects to be correlated with the independent variables (Mundlak, 1978). Thus, our general advice is to use "fixed-effects" and not "random-effects" multilevel MRAs.[29] In the case of minimum-wage research, there is no practical consequence to a preference for fixed-effects MRA. The MRA results for FEML and REML are identical with respect to direction and significance of all moderator variables. As we discussed above, the FEML and REML differ from the WLS and robust MRAs in finding that the inclusion of the unemployment rate is correlated with selection against an adverse employment effect, but this effect is not robust. In all other ways, FEML is practically identical with both the WLS-MRA and the REML-MRA. Most importantly, our central findings are robust to plausible variations in MRA model specifications.

We recommend that meta-analysts focus on those results that are consistent across the multiple WLS, FEML, and cluster-robust MRAs along with the simple FAT-PET-PEESE-MRAs. In our experience, all of these MRA models give consistent results with respect to the existence of publication selection and a genuine effect beyond publication bias. As far as explaining the large variation observed among reported research results, the consistently significant moderator variables identified across multiple WLS, FEML and cluster-robust MRAs should be regarded as important revealed research dimensions. Lastly, a successful meta-analysis will find consistent overall results between the simple FAT-PET-PEESE models and the multiple MRA models regarding the presence of publication selection, the existence of a practically significant empirical effect (or not), and the approximate magnitude of the corrected effect.

5.6 Recap: explaining the heterogeneity of economics research

This chapter discusses approaches to explaining the large variation of empirical economic results routinely reported in the research literature. In our experience, the simple MRA models of publication selection provide an adequate overall estimate of empirical effect corrected for publication bias. However, reviewers and editors will likely demand that any such simple statistics be confirmed by more sophisticated multiple MRAs, which account for potential econometric problems. In addition to minimizing omitted-variable bias, multiple MRAs are needed to understand the large variation of economic and business research findings. Explaining prior research, objectively and comprehensively, is a worthy goal for any research study.

We recommend that meta-analysts take a G-to-S approach to multivariate modeling in order to minimize the potential of identifying spurious research dimensions through data mining (see Figure 5.3).[30] WLS-MRA model (5.6) allows for research dimensions that explain both the reported heterogeneity among results, *Z*-variables, and the propensity that a given finding will be reported and published, *K*-variables.[31] Recall that it is given by

$$t_i = \beta_1 + \sum \delta_j K_{ji} + \beta_0/SE_i + \sum \beta_k Z_{ki}/SE_i + u_i \tag{5.6}$$

where t_i is the reported *t*-value of the *i*th reported effect, and SE_i is the standard error of this effect.

To ensure the robustness of the relevant explanatory research dimensions, a number of alternative MRA model specifications and methods need to be explored and reported (see Figure 5.3). When multiple estimates are typically reported per study, MRA methods that explicitly accommodate potential within-study dependence must be investigated. The best models are cluster-robust and the fixed-effects panel or multilevel (FEML) MRA. We believe random-effects (REML) models will be routinely invalid in meta-analysis due to likely correlation between the unobserved study effects and the moderator variables. But this topic requires further research. If researchers wish to be sure of the validity of their chosen MRA model and thereby to convince reviewers and editors, a Hausman specification test can be used to differentiate between “fixed” and “random” effects, and a Breusch–Pagan Lagrange multiplier test can determine whether a multilevel model is needed in the first place.

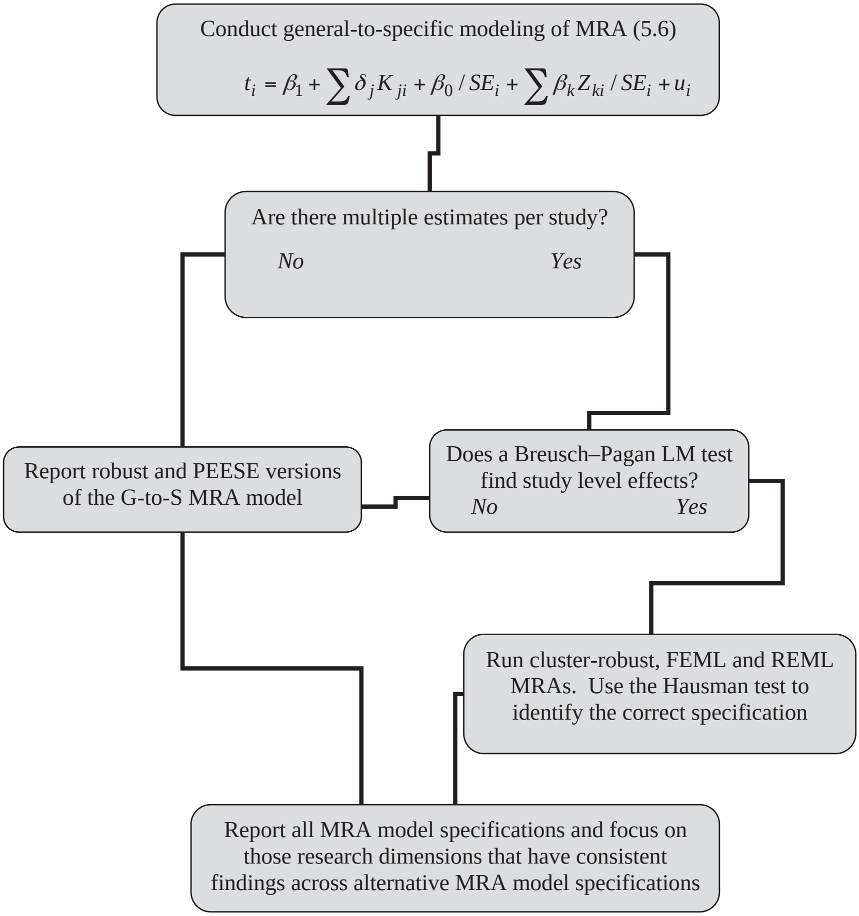


Figure 5.3 Schema for investigating research heterogeneity

A successful meta-analysis is one where the overall findings – in terms of the overall degree of publication bias and the existence of a genuine empirical effect – found by the simple FAT-PET-PEESE-MRAs are consistent with those contained in the more complex multiple MRAs. Those research dimensions consistently identified across different multiple MRA models may be regarded as principal drivers of the reported variation of empirical research results. Robustness is the key characteristic of a genuine understanding of economic research. In the following chapter, we present a more formal, theoretical and technical discussion of the MRA models.

6 Econometric theory and meta-regression analysis

Thus far, we have introduced and applied the conventional battery of statistical methods and approaches to meta-regression analysis. The purpose of this chapter is to delve a bit deeper into the statistical foundation of MRA. This book is intended to be a practical guide for researchers who wish to conduct meta-analyses in economics and business; thus, we have avoided non-essential mathematics and associated technicalities. In this chapter, we present: a theory of MRA directly derived from econometric and statistical theory; a more mathematical and detailed representation of the typical MRA models needed for economic and business applications; a more careful derivation of our MRA models of publication selection; and further technical details about the use of panel (or multilevel) methods in meta-analysis. In the process, we will show how MRA results are entirely unaffected by issues of observed and unobservable study quality when properly modeled. Applied researchers may wish to skim the cream of this chapter.

6.1 The theory of meta-regression analysis

The theory of MRA is firmly established by statistical theory. For empirical economics, econometric theory mathematically and rigorously derives the properties and distribution of statistical estimates, which are typically regression coefficients. Because reported statistical estimates are the dependent phenomenon of MRA, the statistical properties of these econometric estimates entail a structure, hence *theory*, for the associated analysis.

Conventional economic theory typically begins with some generic objective function and derives general behavioral relations from the first-order conditions of the associated optimization problem. However, these derived economic relations rarely specify particular functional forms of the key economic relations, and they are also silent about the required random errors. Arguably, the most important part of an empirical economic model concerns the properties of these random errors such as whether they are independently and identically distributed (i.i.d.) and uncorrelated with the explanatory variables. To complete the necessary specification of an empirical economic relation, conventional practice assumes arbitrary functional forms and tacks on *ad hoc* error terms that possess the needed statistical properties without referring to the underpinning economic theory.

In contrast, econometric theory derives distributional properties of empirical estimates from weak assumptions about the structure of the data used, the underlying relationship, and their connections to the unobserved random errors. When applied econometricians report any statistics, say the *t*-value, for a given empirical estimate, they have implicitly or explicitly made all of the necessary assumptions about the structure of the underlying economic relationship and the error terms.[1] Otherwise, the applied researchers' reported estimate would have unknown properties, and their reported statistics (*t*-values, *p*-values, *etc.*) would be invalid. Taking applied econometric research at face value implies, at a minimum, that the asymptotic distribution of a reported estimate is known and well behaved. From the perspective of meta-analysis, applied empirical work (our meta-data) is assumed to represent what it claims, much as applied econometricians assume that their data represents what the associated governmental agency, which collected the data, purports. Of course, it is the responsibility of the meta-analyst, like the econometrician, to identify important errors or omissions in their data and to accommodate those deficiencies when possible. However, to make progress, all empirical researchers must first assume that their data are valid, at least provisionally. With data validity as the null hypothesis, deviations from the ideal may be carefully traced, modeled and empirically tested. Tracking these deviations from ideal econometric properties provides much of the structure of MRA.

Recall from elementary econometrics that regression estimates, $\hat{\alpha}_k$, will have an asymptotic normal distribution under very general and weak conditions.[2] We have:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{u} \qquad \text{and} \qquad \hat{\boldsymbol{\alpha}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \tag{6.1}$$

where $\hat{\boldsymbol{\alpha}}$ is a $K \times 1$ vector of estimated regression coefficients, $\hat{\alpha}_k$, $\mathbf{X}^t$ is the transpose of an $n \times K$ matrix, $\boldsymbol{X}$, of exogenous explanatory variables, $\boldsymbol{Y}$ is the economic phenomenon investigated and $\mathbf{u}$ is a vector of random errors. As long as the errors are i.i.d., $\mathbf{X}$ is of full rank, and $p\text{lim}\ (n - 1)\mathbf{X}^t\mathbf{X}$ is also of full rank, then a law of large numbers will ensure that $\hat{\boldsymbol{\alpha}}$ will be consistent (asymptotically unbiased) and have an asymptotic normal distribution (Davidson, 2000). In practical samples, the *t*-distribution usually gives an acceptable approximation. These widely applicable econometric properties establish known distributions for MRA's dependent variable, at least in large samples.[3]

However, the potential econometric violations of this simple and clear picture are legion. The vast majority of econometrics concerns complications, exceptions and weaknesses of the assumed statistical properties of econometric estimates (equation (6.1)). Econometric theory and practice clearly map the many weaknesses of conventional econometric theory for specific applications and difficulties. This map identifies relevant moderator variables for the MRA model and thereby gives structure to the resulting multiple MRAs. Exceptions and complications to the simple econometric story of well-behaved estimates will give theoretical structure to our MRAs.

In MRA, our dependent variable is often an estimated regression coefficient (say, $\hat{\alpha}_1$), and it will be asymptotically normal with estimated variance SE^2.

Furthermore, $(\hat{\alpha}_1 - \alpha_1)/S_{\hat{\alpha}1}$ has a *t*-distribution (Davidson and MacKinnon, 2004: 140–1). Otherwise, the reported econometric research results would be invalid, not representing what applied researchers claim. In practice, the *t*-distribution is likely to be a good approximation as long as the residuals are not highly skewed. In the empirical literature, the ratios of regression estimates to their standard errors are almost always assumed to have a *t*-distribution under the null hypothesis that $\alpha_1 = 0$. In any case, these coefficients will be asymptotically normal. For the purposes of meta-analysis, it is sufficient that the sampling distributions of the empirical estimates are approximately symmetric and that the estimated regression coefficients divided by their standard errors to be approximately *t*-distributed under the null hypothesis that $\alpha_1 = 0$. These properties of estimated regression coefficients are sufficient to establish the independence and asymptotic normality of MRA errors and hence the validity of MRA estimation and hypothesis testing – see equations (6.2)–(6.4) and the discussion below.

These statistical properties of regression coefficients also imply that a funnel graph will be symmetric. Funnel symmetry requires only that the regression estimates be symmetrically distributed around the true effect, α_1, and independent of their standard errors. Both of these conditions follow directly from the fact that $(\hat{\alpha}_1 - \alpha_1)/S_{\hat{\alpha}1}$ has a *t*-distribution. If the magnitude of an estimated effect is independent of its standard error, and hence precision, then there will be no pattern to the funnel graph other than predictable heteroskedasticity.[4] When regression estimates are independent of their standard errors, only *random* sampling errors cause estimates to vary for any given level of precision (or standard error). Importantly, the independence of $\hat{\alpha}_1 - \alpha_1$ from the standard error of $\hat{\alpha}_1$ is a well-known property of the *t*-distribution (Davidson and MacKinnon, 2004: 140–1), as well as its symmetry. Thus, the symmetry of the funnel graph, centered on α_1, derives from the well-known and widely assumed statistical properties of regression estimators.

A potential exception to this simple derivation of a symmetric funnel might occur if some estimates contain systematic bias, or equivalently sample a population with a different underlying empirical effect. This exception means that the set of reported estimates will have heterogeneity. Explaining this heterogeneity is the central role of multiple MRA (recall Chapter 5), and all meta-analyses in economics will need to explicitly model potential heterogeneity using multiple MRA. Meta-analysts must always control for likely disparities from such a simple, unbiased view of the reported empirical estimates. These complications are explicitly addressed by multiple MRAs and discussed further below.

A second potential exception to funnel symmetry resulting directly from conventional econometric theory is small-sample biases. In some applications (e.g. estimating the regression coefficient of a lagged dependent variable) the reported estimates are known to have small-sample biases but nonetheless remain consistent. In such cases, the source of a funnel's asymmetry cannot be unambiguously identified as either publication selection or small-sample bias. This dilemma is recognized and discussed in Stanley (2004). For practical purposes, however, this issue is largely irrelevant. Because both publication selection and small-sample bias decrease with sample size and therefore with precision, the publication selection correction methods introduced in Chapter 4 can track and correct for both types of bias. Obviously, we seek to minimize all biases in a research literature, regardless of their source. The only remaining ambiguity is whether to label the identified bias, "publication selection" or "small-sample."

A skeptic might still point out that this approach is naïve because it assumes that the reported empirical estimates have desirable statistical properties that they are widely known not to possess in many actual applications. A lake of ink has been used to document how various econometric problems (omitted-variable bias, simultaneity bias, incorrect functional forms, heteroskedasticity, nonstationarity, etc.) often invalidate the reported statistics by altering their expected values, consistency, and/or variance–covariance matrix. However, this observation is nothing new to this book or to MRA. Recall that the estimation of these misspecification biases was precisely the motivation for developing MRA in the first place (Stanley and Jarrell, 1989).

The nice thing about misspecification bias is that, by definition, it adds a term to the expected value of a reported estimate. For example, omitted-variable bias adds a term, $\gamma_1\alpha_2$, to the expected value of the estimated regression coefficient, $\hat{\alpha}_1$. When variable X_2 is omitted from an applied econometric model, $E(\hat{\alpha}_1) = \alpha_1 + \gamma_1\alpha_2$. Note how the magnitude of this bias, $\gamma_1\alpha_2$, is the product of population parameters, where α_2 is the regression coefficient of the omitted variable, X_2, and γ_1 is the regression coefficient from an auxiliary regression of X_2 on X_1. Both γ_1 and α_2 represent population values of regression coefficients, which will be independent of other variables and research dimensions. Such a shift of expected values can be represented by a dummy variable to denote the omission of a relevant variable, which, in turn, enables the MRA to estimate $\gamma_1\alpha_2$, or its equivalent, when other types of misspecification biases are considered. Other forms of misspecification will also cause additive biases, and their presence should be identified by other moderator variables, which may also be included in the MRA.[5] The well-established and widely known structure of econometric misspecification biases provides a theoretical basis for MRA models.

To understand the typical properties of reported empirical estimates, the resulting meta-regression models used to explain them and potential deviations from these simple cases, we turn to a more formal representation of meta-regression models. Suppose we are interested in summarizing and explaining the observed variation of some estimated regression coefficient (perhaps, an elasticity), e_i. The basic form of the meta-regression model is

$$\mathbf{e} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{6.2}$$

Here $\mathbf{e}$ is an $L \times 1$ vector of all the reported empirical effects in an empirical literature of L estimates, which are often regression coefficients, $\hat{\alpha}_{1i}$.[6] $\mathbf{M}$ is an $L \times K$ matrix of moderator variables, the first column of which contains 1s. In Chapter 5, we grouped moderator variables into Z- and K-variables. Here, for the sake of simplicity, $\mathbf{M}$ can be either or both. $\boldsymbol{\beta}$ is a $K \times 1$ vector of MRA coefficients, the first of which represents the "true" underlying empirical effect investigated.[7] $\boldsymbol{\varepsilon}$ is an $L \times 1$ vector of residuals representing the estimation errors of the reported empirical effects. Recall from previous discussions that the moderator variables will include dummy variables that allow for any likely misspecification or selection bias. Likely heterogeneity and potential violations of the symmetry of reported effects are explicitly modeled by $\mathbf{M}$ in (6.2).

Let us return to the theory of this MRA model (6.2). Like all regression models, including conventional econometrics, the entire theoretical structure is contained in two substantively distinct components: the random error terms ($\boldsymbol{\varepsilon}$) and the explanatory, deterministic structure ($\mathbf{M}\boldsymbol{\beta}$). The proper structure of these errors is critical for reliable estimation. Unlike conventional applied economics, MRA does not require additional *ad hoc*, atheoretical assumptions about these regression errors. In MRA, ε_i is the estimation error of our targeted empirical finding, and its statistical properties are well known and fully specified in the research literature investigated – advantage meta-analysis.

Next, consider the explanatory deterministic structure of our MRA model, $\mathbf{M}\boldsymbol{\beta}$. Here too, statistical theory (e.g. omitted-variable, publication and other misspecification biases) provides theoretical structure. We know that these biases, by definition, impart additive terms on an expected value of the associated estimate, which is the dependent variable in (6.2). In other cases, economic and measurement theories will give additional structure to an MRA. Further moderators are required when we have theoretical reasons to believe that differences in how estimates are measured or calculated might systematically affect them (e.g. compensated price elasticities or those calculated from alternative demand functions). Practical issues of measurement and data often have important effects on observed research results. Needless to say, such issues also plague conventional applied econometric research.

For example, our greater knowledge of the structure of MRA advises us to use weighted least squares (WLS) in all cases. Unlike conventional econometric regression models, MRA residuals, $\boldsymbol{\varepsilon}$, can never be assumed to be i.i.d., because the standard errors of the reported effects vary widely. That is, meta-analysts directly observe large heteroskedasticity among reported estimates of effects, which define the dependent variable in their meta-analyses. Thus, simple ordinary least squares (OLS) is never the preferred approach for any MRA model, but rather weighted least squares. WLS should always be used, at least for a baseline (Stanley and Jarrell, 1989).[8] The WLS estimate of MRA (6.2) is

$$\hat{\boldsymbol{\beta}} = \mathbf{M}^t\boldsymbol{\Omega}^{-1}\mathbf{M})^{-1}\,\mathbf{M}^t\boldsymbol{\Omega}^{-1}\mathbf{e} \tag{6.3}$$

where:

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_L^2 \end{bmatrix}$$

and σ_i^2 is the variance of the ith estimated effect, e_i, and its sampling error, ε_i (Davidson and MacKinnon, 2004; Green, 1990).

Equation (6.3) is a generalized least squares (GLS) estimator. WLS is a special case of GLS where the variance–covariance matrix, $\mathbf{\Omega}$, has this specific diagonal structure noted above.[9] When the parameters in $\mathbf{\Omega}$ are known, GLS is the best linear unbiased estimator (Green, 1990). More relevant to empirical work is the fact that this approach still has very desirable properties when consistent estimates of σ_i^2 are used in their place. With consistent estimates of σ_i^2, this feasible GLS version of (6.3) will itself be consistent, asymptotically efficient, and asymptotically normal (Wooldridge, 2002: 160–2).[10]

Here too, meta-analysts are in a better position than conventional econometricians. By coding the statistical results of an entire empirical literature, they have ready access to the informative content of n_i observations from each of the L estimates reported in the literature. That is, rather than using estimated squared residuals from (6.2) and some skedastic function as a rough estimate of an individual variance (Davidson and MacKinnon, 2004), each study in the research literature provides a direct estimate of the needed variance, SE_i^2, from the n_i observations used in that study. By the assumptions made in each research study, the square of the standard error of the reported estimate will be a consistent estimate of σ_i^2 and often unbiased as well. Our WLS estimation strategy (6.3) is easily implemented by most statistical packages using analytic weights = $1/SE_i^2$ in a WLS routine.[11]

With L estimates of σ_i^2, SE_i^2, we can also divide MRA model (6.2) by SE_i to get the entirely equivalent WLS-MRA in the form:

$$t_i = (1/SE_i)\mathbf{M}_i\boldsymbol{\beta} + (1/SE_i)\varepsilon_i \tag{6.4}$$

where t_i is the t-value of reported effect i (Davidson and MacKinnon, 2004: 261; Wooldridge, 2002). Estimating MRA (6.4) by OLS is equivalent to the feasible GLS that uses SE_i^2 to estimate σ_i^2 in (6.3).

This form of the WLS-MRA, equation (6.4), makes the role of precision ($1/SE_i$) quite clear. However, some applied researchers have difficulty interpreting the MRA coefficients from (6.4) correctly. Thus, in applications, we recommend estimating WLS-MRA model (6.2), which calculates (6.3), and specifying $1/SE_i^2$ as the analytic weights in standard statistical packages. Recall our previous discussions of MRA models (4.1) vs. (4.2) and (5.5) vs. (5.6).

Hedges and Olkin (1985: 174) and Konstantopoulos and Hedges (2004: 293) have argued that the standard errors of the WLS statistical packages are wrong for meta-analysis and need to be divided by the square root of the mean square error (MSE). This point has been repeated by many other meta-analysts. We do not agree. What is at issue is whether σ^2 in the variance–covariance matrix, $\sigma^2(\mathbf{M}^t\mathbf{\Omega}^{-1}\mathbf{M})^{-1}$, must be constrained to be equal to 1 or not. If we assume that there is no between-study heterogeneity and that σ_i^2 fully reflects the uncertainty of each individual estimated effect, then the WLS variance–covariance matrix does reduce to $(\mathbf{M}^t\mathbf{\Omega}^{-1}\mathbf{M})^{-1}$ and $\sigma^2 = 1$, as Hedges and Olkin (1985) and Konstantopoulos and Hedges (2004) suggest. Their point is technically correct, when the "fixed-effects" model is used and we assume that there is no between-study heterogeneity ($\boldsymbol{\tau}^2 = 0$ in conventional meta-analysis notation). However, in economics research, we have not seen a case where there is no excess heterogeneity and therefore see no need to constrain $\sigma^2 = 1$. Allowing the research record to determine the best value of σ^2 permits the WLS standard errors to accommodate this overall heterogeneity. Forcing $\sigma^2 = 1$, as implied by Hedges and Olkin's (1985) recommendation, typically reduces the size of the WLS confidence intervals, making these estimates seem more precise and their *t*-values larger. But with excess heterogeneity, these "fixed-effects" WLS coefficients will likely have more variation than these formulas suggests. This is the central weakness of "fixed-effects" MRA compared to "random-effects" MRA; that is, "fixed-effects" MRAs tend to report standard errors that are too small for the actual uncertainty involved. Thus, we see no reason to make this weakness worse by overriding standard WLS reported results by forcing σ^2 to be 1. Our view is to allow standard WLS packages to use the data to determine σ^2 and to compensate for at least some of the likely excess heterogeneity. See the Appendix to this chapter for a further discussion of this issue.

The weakness of our MRA model resides in the case where the original regression models (6.1) are misspecified in a manner that biases the estimated variances. Recall that the design matrix, **M**, accounts for potential biases, but it does not further correct for inconsistent estimates of the standard errors. An alternative approach for potential inconsistencies in estimating the standard errors is to use a conventional feasible GLS estimate of the WLS-MRA given in (6.3); that is, one that uses the estimated individual residuals to estimate σ_i^2. The real possibility that SE_i^2 may be biased suggests that we should be conservative in calculating MRA standard errors. We recommend that meta-analysts use heteroskedasticity-robust and/or cluster-robust standard errors in conjunction with MRA model (6.4) as additional insurance.[12]

This statistical theory of MRA has been corroborated in simulations where dummy variables are used in MRAs to estimate and accommodate misspecification biases as represented by **M** in MRA model (6.2) (Koetse *et al*., 2010). In these simulations, both random unobserved effects on individual estimates and regression misspecifications were introduced, and yet the WLS-MRA model (6.4) that uses dummy variables to identify the presence of potential misspecification bias outperformed a mixed-effects estimator that explicitly allows for individual random effects. Our WLS-MRA is found to do a remarkable job in estimating both the misspecification bias and the underlying true empirical effect.[13] Thus, MRA has its foundation in well-established econometric and statistical theory. Because MRA models are derived from statistical theory, they can easily be corroborated and modified, when necessary, by Monte Carlo experiments.

6.2 Improving meta-regression analysis with unbalanced panel models

Although it is econometric theory that imbues meta-regression with its theoretical structure, more practical econometrics provides further specification of MRA models. In particular, when multiple estimates of some economic phenomenon are reported by research studies, this fact alone offers meta-analysts a great opportunity to improve their MRA estimates and thereby more accurately depict research. When there are several estimates per study, they may jointly be influenced by some common unreported or unobservable factor: the quality of the study, the ideology of the researcher, the authors' funding source, or even some unique interpretation (or misunderstanding) of the associated economic and econometric theories. Regardless, such a multiple estimate research structure induces potential dependence among the reported estimates in each study, and this dependence must be addressed to ensure the validity of the MRA results.

Because referees and editors often demand robustness checks for any empirical finding, a multiple-estimate research structure is quite common in empirical economics. However, only one of our selected examples, the employment effect of the minimum wage, has the full multidimensional data structure required for panel analysis. Recall that we found 1,474 estimates of the employment effect of raising the US minimum wage in 64 studies.

In Chapters 4 and 5, we reported the unbalanced-panel (or multilevel) findings for the minimum-wage research. In the case of minimum wages, the MRA results from pooled ordinary least squares (POLS) are very similar to what the more sophisticated panel methods find. However, this consistency need not be the case, and when there are differences, econometric theory clearly favors panel methods. The purpose of this section is to present a more formal unbalanced panel model for MRA and to explore in greater detail its implications for business and economics research.

By including study-level effects, S_s, in our previous MRA model (6.2), we can accommodate potential dependence among estimates within a given study:[14]

$$e_{is} = \beta_0 + \sum_{j=1} \beta_j M_{jis} + S_s + \varepsilon_{is}, \qquad i = 1, 2, ..., m_s, s = 1, 2, ..., K \tag{6.5}$$

Here m_s is the number of estimates in study s, and K is the number of studies. This MRA model has an "unbalanced" panel structure because m_s varies across studies. Although econometricians are most familiar with panels that are pooled time-series and cross-sectional data, any multidimensional data structure may be regarded as a panel. Rosenberger and Loomis (2000b) were the first to recognize that the typical data structure encountered in meta-econometrics may be interpreted and analyzed as an unbalanced panel. MRA model (6.5) can be estimated using either "fixed" or "random" effects panel or multilevel methods.[15]

6.2.1 Fixed vs. random-effect panel MRAs

There is considerable misunderstanding about the meaning of "fixed" vs. "random" effect panel methods. According to Wooldridge's (2002) view, all panel models are "random," and the conventional distinctions between them are just "wrongheaded":

> In modern parlance, "random-effect" is synonymous with zero correlations between the observed explanatory variables and the unobserved effects. ... [T]he term "fixed-effect" does not usually mean that [S_s] is being treated as non-random; rather it means that one is allowing for arbitrary correlation between the unobserved effect [S_s] and the observed explanatory variables [M_{jis}].
>
> (Wooldridge, 2002: 252)

"Fixed-effects" panel methods can be considered the more general approach that allows for correlation between the study-level effects and the moderator variables. To better understand why "fixed-effects" methods are more general and robust, we first consider the "fixed-effects" approach to estimation. MRA model (6.5) can be estimated by a "fixed-effects" panel model in two equivalent ways, both of which use OLS. The most obvious is to replace S_s by K dummy variables, $\sum_{s=1}^{K}\delta_s D_{is}$, assuming that one omits the intercept. This least-squares dummy variable (LSDV) approach also allows us to $1/SE_i^2$ as the analytic weights.[16] A second equivalent approach to "fixed-effect" panel estimation subtracts study averages from all observed values (Wooldridge, 2002: 267):

$$e_{is}^d = \sum_{j=1} \beta_j M_{jis}^d + \varepsilon_{is}^d, \qquad i = 1, 2, ..., m_s,\ s = 1, 2, ..., K \tag{6.6}$$

where $e_{is}^d = e_{is} - \overline{e}_s$, $M_{jis}^d = M_{jis} - \overline{M}_{js}$, and the bar variables, $\overline{e}_s$ and $\overline{M}_s$, are the sth study average of the reported effect and moderator variables, respectively. Note that the study-level effects, S_s, disappear from this model entirely, because S_s is constant within each study. Moderator variables that do not vary at all within studies will also drop out, and their effects cannot be estimated by fixed-effects panel methods. Subtracting the study average of S_s makes each difference ($S_s - \overline{S}_s$) equal to zero. No part of S_s will be contained in the error terms; hence correlation of S_s with the moderator variables causes no bias or inconsistency. Further, note that all influences from any observed or unobservable variable that is constant for each study, such as study quality, is entirely eliminated by this model. This fact has important implications for the quality of our MRA inferences, which are explored further in Section 6.2.2.

In conjunction with panel models, the meta-analyst should also use cluster-robust standard errors and the WLS multiple MRA(equation (5.6)). Unlike conventional econometric panels, it is very unlikely that estimates within studies will exhibit the type of dependence routinely seen in time series. However, the variance within studies might well differ from study to study even after the systematic variation is fully accounted for. Thus, it is prudent to use cluster-robust standard errors.

In contrast, "random-effects" panel methods replace S_s with a random effect, v_s, and a feasible GLS strategy is employed to estimate a rather complex variance–covariance matrix, $\mathbf{\Omega}$ (Wooldridge, 2002).[17] However, one must further assume that the study-level effects, S_s, and moderator variables, M_{jis}, are independent, if the resulting estimates are to be consistent and unbiased. If S_s and M_{jis} are correlated, "random-effects" panel estimates become biased. To see this, define a new composite error term, $v_{is} = v_s + \varepsilon_{is}$, and (6.5) becomes:

$$e_{is} = \beta_0 + \sum_{j=1} \beta_j M_{jis} + v_{is}, \qquad i = 1, 2, ..., m_s,\ s = 1, 2, ..., K \tag{6.7}$$

Note that if M_{jis} is correlated with v_s it will also be correlated with the composite error term in (6.7), v_{is}. As discussed in every econometrics textbook, whenever the independent variable of a regression model is correlated with the regression's error terms, estimates will be biased and inconsistent. Essentially, the overlap between the random and deterministic components of the regression model does not permit a clean separation or estimation.

Because there is no reason to rule out correlations between moderator variables and study effects in a meta-regression context, meta-analysts should use "random-effects" unbalanced panel models with great caution. Conventional "fixed-effects" panel methods are the more general and robust approach, thus they earn the preferential position unless there are good reasons to the contrary.[18] Or, as discussed in Chapter 5, one can use the Hausman test to test for this correlation and thereby to choose between fixed- and random-effects models (recall Figure 5.3). However, we have further reasons to suspect widespread correlations between study-level effects and moderator variables in MRA. In most cases, unobservable study quality will be correlated with observed and coded research choices of methods and variables, which will induce bias if a "random-effects" model is used.

6.2.2 A note on study quality

For decades, labor economists have called for the collection and greater use of longitudinal (or panel) data as a way to accommodate unobservable ability. For example, the omission of worker ability in the conventional log-wage equation has long been recognized as a bias in the estimated returns to education (Griliches, 1977). Because education and unobserved ability are almost certainly positively correlated, the omission of ability biases the estimated returns to education. To cope with this serious problem, labor economists use proxies for ability, instrumental variables, and panel methods when longitudinal data are available. Fortunately, longitudinal data sources have long been available, such as the panel study of income dynamics and the national longitudinal survey of youths. With such data, panel methods can do much to correct for the bias of omitting worker ability, assuming that ability does not change much over time.

In a meta-analysis context, study quality is likely to play a similar role as worker ability, and like ability, study quality is probably correlated with important explanatory variables, potentially biasing any MRA results. A contaminating correlation is likely to exist between unobserved study quality and observed methodological and model specification choices. That is, study effects, S_s, are likely to reflect study quality, at least in part, and study quality should be correlated with some of the moderator variables that code for the types of econometric techniques, models, and data used, the precision of the estimate, and the important variables omitted from the original regression relation. Thus, the omission of study quality could bias MRA coefficients as the OLS estimates of the returns to education were biased.

Within-study differences in econometric techniques, omitted variables, precision, etc., by definition, cannot measure study-level quality. They are also *observable* and should, therefore, be explicitly included in the MRA.[19]

Fortunately, fixed-effect panel methods can render unobserved *study* quality harmless. As discussed in Section 6.2.1, any study-level effect can be filtered out of the model, ensuring that the rest of the model can be adequately estimated. To demonstrate this, return to our MRA model (6.5) but add a study quality variable, Q_s:

$$e_{is} = \beta_0 + \sum_{j=1} \beta_j M_{jis} + S_s + \gamma Q_s + \varepsilon_{is}, \qquad i = 1, 2, ..., m_s, s = 1, 2, ..., K \tag{6.8}$$

Here, we expect that Q_s will be correlated with some of the moderators, M_{jis}, perhaps highly so. After all, if "study quality" is to deserve this name, it should be correlated with, for example, the precision of the reported empirical effects, the choices of econometric models that researchers make or whether researchers fail to include obvious important variables into their models.

Without loss of generality, we can embed unobserved study quality into the study effect. Define $v_s = S_s + \gamma Q_s$, and replace this in (6.8):

$$e_{is} = \beta_0 + \sum_{j=1} \beta_j M_{jis} + v_s + \varepsilon_{is}, \qquad i = 1, 2, ..., m_s, s = 1, 2, ..., K \tag{6.9}$$

This MRA model is now identical to our previous equation (6.5), and fixed-effects panel methods can consistently and unbiasedly estimate the MRA regression coefficients by entirely filtering out the study-level effects, whether v_s or S_s. Recall that fixed-effects panel methods work even when these study-level effects (including quality here) are correlated with the included moderator variables. However, this is not true for "random-effects" panel methods or with pooled OLS. If OLS is used to estimate (6.8) but Q_s cannot be observed, then we have the classic case of omitted-variable bias. Only fixed-effects panel models can fully avoid the potential bias from omitting study quality.[20]

This brief note is meant to complement our previous discussion of research quality in Chapter 2. It is our view that objective and observable dimensions of study quality should be coded and included in MRA. Potentially more pernicious are those aspects of research quality that are more difficult or impossible to measure objectively. Nonetheless, we find it quite comforting that such unobservable factors will not contaminate MRA panel estimates, even under the worst circumstances where they are highly correlated with the estimated MRA effects.

It may also be worth pointing out that these desirable properties of panel MRAs hold for a broader class of unobservable and "unmentionable" study effects such as researcher ideology, research funding source, their institutional affiliation and the strength of the commitment that researchers have for a given theory. Because such potentially contaminating influences do not vary across estimates within a study, they are swept up into the study effect, v_s in (6.9). Regardless of the strength of the influence that such factors might have on research, like study quality, they can be folded into the study effects, and fixed-effect panel methods can estimate the remaining observable and objective dimensions unbiasedly. Even the most "politically sensitive" issues of economic and business research can be accommodated without contaminating the remaining MRA estimates when multiple estimates are reported per study.

6.3 Meta-regression models of publication selection

In previous chapters, we introduced simple meta-regression models of publication selection, where the reported empirical effect is some function of its standard error. The simplest of these models is the FAT-PET-MRA. Recall that:

$$effect_i = \beta_0 + \beta_1 SE_i + \varepsilon_i \tag{4.1}$$

This meta-regression of an estimate and its standard error was first introduced by Egger *et al.* (1997) to serve as a test for the presence of publication bias. This Egger meta-regression is itself a generalization of the well-known Galbraith diagrams (Galbraith, 1988) that adds a constant term to the WLS version of (4.1):

$$t_i = \beta_1 + \beta_0(1/SE_i) + v_i \tag{4.2}$$

In spite of the intuitive appeal of these relations, they lack a rigorous statistical foundation. Although it seems apparent that smaller studies with correspondingly larger standard errors would need to search harder over different data subsets, alternative specifications and methods in order to achieve statistical significance when the underlying phenomenon is small or non-existent, this connection between a reported estimate and its standard error begs for a more rigorous grounding. The purpose of this section is to provide a mathematical argument for this dependence of a reported estimate and its standard error when there is publication selection for statistical significance. In the process, we hope to provide a better understanding of this relation and its limitations.

Publication selection is analogous to the better-known bias that arises through sample selection, famously addressed by Heckman (1979).[21] Take the example of gender wage discrimination. The gender wage gap is estimated by comparing the estimated returns to worker productivity measures from samples of male and female worker wages (e.g. Oaxaca, 1973; Jacobsen, 1994). A problem arises because wages are only observed for employed workers, those who have reservation wages lower than the observed market wage rate. The decision to participate in the labor market is itself a function of wages, but wages might be differentially affected by gender discrimination. Therefore, a regression on just employed workers may provide biased regression estimates. Observing a worker's wages and discrimination may be endogenously related. To address this issue, labor economists have long employed a Heckman correction for this sample selection bias, and a meta-regression of the gender wage gap finds that using a Heckman correction greatly increases (by approximately 18 percentage points) the reported gender wage discrimination (Stanley and Jarrell, 1998).

Publication selection involves a similar case of incidental truncation. It is "incidental truncation" because the magnitude of the reported effect (like worker wage) is not directly selected but rather some other variable, the estimate's *t*-value (labor market participation); see Wooldridge (2002: 552). Incidental truncation differs from censored sampling where the dependent variable is itself selected and there is data on the independent variables for both the selected and the unreported samples (Heckman, 1979; Wooldridge, 2002: 552).

With publication selection for directional statistical significance, we observe an estimated effect only if $effect_i/SE_i > a$, where a is the critical value of the standard normal distribution. By referring to the well-known conditional expectation of a truncated normal distribution, it is easy to show that observed effects will depend on the population (or "true") effect plus a term that reflects the selection bias, which is equal to the standard error times the inverse Mills ratio:

$$E(effect_i \mid truncation) = \alpha_1 + \sigma_i \cdot \lambda(c) \tag{6.10}$$

where $\lambda(c)$ is the inverse Mills ratio, α_1 is the "true" effect, which is the expected value of the original distribution, σ_i is the standard error of the estimated effect, and $c = a - \alpha_1/\sigma_i$. Because $effect_i$ is asymptotically normal with mean α_1 and standard deviation σ_i, equation (6.10) follows directly from Theorem 21.2 of Green (1990) (see also Johnson and Kotz, 1970). Relation (6.10) has the same general form as a Heckman regression (Davidson and MacKinnon, 2004: 488).

When we replace sample estimates for the population values in (6.10) we get

$$effect_i = \alpha_1 + SE_i \cdot \lambda(c) + \varepsilon_i \tag{6.11}$$

If one further assumes that the inverse Mills ratio is constant, we have our FAT-PET-MRA equation (4.1). Like (6.11), the more familiar Heckman regression adds a term containing the inverse Mills ratio and σ_i. Thus, statistical models of truncation and selection offer a simple meta-regression relation between observed effect and its standard errors, giving us a rigorous foundation for the publication selection methods described and applied in Chapter 4.

However, the linear MRA defined by equation (4.1) further assumes that $\lambda(c)$ is approximately constant with respect to σ_i. Unfortunately, we know better. In general, $\lambda(c)$ is not constant, and variations of $\lambda(c)$ can cause the MRA estimate of the mean of the full distribution of effects, $\hat{\beta}_0$, to be biased and inconsistent. This complication causes considerable difficulty in finding an unbiased and consistent corrected estimate of the empirical effect in question.

To understand this problem in context, consider how the conventional correction for sample selection works. When we have data on the explanatory variables for both the selected and unreported samples, the conventional Heckman regression consistently estimates the corrected effect using a two-step method, where the first step models the probability of being selected, and in the second step, estimates from the selection equation are used to calculate the inverse Mills ratio in a Heckman regression. In effect, the estimated selection relation gives a sample estimate of the selection bias term, $\sigma_i \cdot \lambda(c)$.[22] To state the obvious, this conventional approach is not available to the meta-analyst, because we have no information on the characteristics of the unreported values; thus, no direct way to model the selection process or to estimate the Heckman regression.

So what is the "second" best strategy? Somehow we need to estimate the publication bias term, $\sigma_i \cdot \lambda(c)$, however crudely, using information contained only in reported research findings. Further, we know that $\lambda(c)$ is itself a function of σ_i, but, unfortunately, it is not a simple function of σ_i.

To see what type of relation we must approximate, take the derivative of (6.10) with respect to σ_i:

$$\begin{aligned} \partial E(effect_i \mid truncation)\, \partial\sigma_i &= \lambda(c) + \sigma_i \cdot \partial\lambda(c)/\partial\sigma_i \\ &= \lambda(c) + \sigma_i \cdot (\partial\lambda(c)/\partial c) \cdot (\partial c/\partial\sigma_i) \end{aligned} \tag{6.12}$$

However, Heckman (1979: 159) shows that $\partial\lambda(c)/\partial c = \lambda(c)^2 - c\lambda(c)$, which gives

$$\partial E(effect_i \mid truncation)\, \partial\sigma_i = \lambda(c) + (\alpha_1/\sigma_i) \cdot (\lambda(c)^2 - c\lambda(c)) \tag{6.13}$$

In general, this expression is a rather complex non-linear function of σ_i; thus, some rough approximation such as the power series will need to be employed to estimate the expected empirical relation between a reported estimate and its standard error. Using a power series to approximate this conditional expectation is the starting point of PEESE-MRA model (4.3):

$$effect_i = \beta_0 + \beta_1 SE_i + \beta_2 SE_i^2 + \varepsilon_i \tag{6.14}$$

Inspection of the limiting relations suggests that the bottom of this parabola should occur when SE_i =0 (recall Section 4.3.4 and Box 4.8). Constraining a second-order power series to have its perigee at $SE_i = 0$ implies that $\beta_1 = 0$, removes the linear term from (6.14), and gives the PEESE-MRA (4.4):

$$effect_i = \beta_0 + \beta_2 SE_i^2 + \varepsilon_i \tag{4.4}$$

Two separate simulation studies have confirmed the viability of using $\hat{\beta}_0$ from the WLS version of (4.4) as a corrected estimate of empirical effect. Stanley and Doucouliagos (2011) compare simple and weighted averages (recall FEE and REE from Chapter 3) to both the linear and quadratic FAT-PET-PEESE-MRAs and find that PEESE has the smallest bias and MSE when there is a genuine empirical effect. These simulations also show that quadratic or cubic power series that are not constrained to have $\beta_1 = 0$ have large bias and MSEs. Unconstrained power series are clearly dominated by PEESE (4.4). Secondly, a team of medical researchers report a "comprehensive simulation study" on 14 different approaches, including "trim-and-fill," to estimating effect when there might be publication bias (Moreno *et al.*, 2009a). Their simulations find no better approach to publication bias correction than $\hat{\beta}_0$ on a combination of four criteria (bias, MSE, variance and coverage percentage).[23] When Moreno *et al.* (2009b) apply these publication correction methods to randomized clinical trials of antidepressants, they use PEESE and found it to be the best way to correct for publication selection bias.

There is an important special case for the relation between the expected value of a reported effect and its standard error that must be mentioned. When the underlying empirical effect is zero, $\beta_0 = 0$, equation (6.13) simplifies to $\lambda(c)$. Recall that $c = a - \beta_0/\sigma_i$, and $\partial E(effect_i \mid truncation) / \partial\sigma_i$ reduces to the inverse Mills ratio evaluated at critical value of the standard normal distribution, *a*, which, of course, is just a constant. Thus, when there is no genuine empirical effect, the expected reported effect will be a multiple of its standard error, and the linear MRA model used in Chapter 4, MRA (4.1), will be correct.[24] This observation is important because it further validates the precision-effect test (H_0: $\beta_0 = 0$), which tests for the presence of a genuine underlying empirical effect beyond publication selection bias. The PET's null hypothesis assumes that $\beta_0 = 0$; thus, the FAT-PET-MRA is correctly specified as a linear relation for testing whether there is a genuine non-zero empirical effect. As a result of this special case, simulations further confirm that the PET estimate from (4.1) is superior to PEESE (4.4) when we accept H_0: $\beta_0 = 0$, but that PEESE (4.4) is statistically more accurate when this hypothesis is rejected (Stanley and Doucouliagos, 2011). As a result, one should only use the PEESE correction if there is first evidence of some genuine effect (i.e. reject H_0: $\beta_0 = 0$).

6.4 In defense of simple statistical methods

Although Chapter 5 and this chapter focus largely on complex and rigorous statistical methods of meta-analysis, we advocate very simple statistical approaches whenever possible. Our experience suggests that simple meta-analytic methods are usually adequate to summarize a research literature. Often these simple statistics are more revealing than sophisticated multivariate analyses. Nonetheless, we acknowledge the need to also employ more complex and econometrically rigorous methods, if for no other reason than to be sure that simple findings are robust. Below we argue that very simple meta-analytic techniques should be reported and that they might possibly do a better job summarizing an empirical literature than more rigorous and complex methods.

To take an odd but revealing example, in a recent *American Statistician* article we demonstrate how it might be better to throw out 90 percent of the research literature and just average the rest (Stanley *et al.*, 2010). Simulations also show that this *top 10* estimator compares well to simple and weighted averages (FEE and REE) and to $\hat{\beta}_0$ from the linear MRA model (4.1). The secret is that the 10 percent of the research that is retained are those estimates that are the most precise – *top 10*. Our *top 10* estimator is not meant to offer a genuine applied approach to meta-analysis but only as a statistical paradox that highlights the seriousness of publication selection. Nonetheless, it also demonstrates how the intelligent use of the simplest statistical methods (a mean) can be more enlightening and statistically valid than the mechanical use of seemingly more rigorous and efficient estimators (such as REE).

6.4.1 Selected heterogeneity and the simple FAT-PET-PEESE-MRA

A sensible case can be made for the preference of the estimates from the simple MRA models (4.1) and (4.4) over more complex multiple MRA models. How is this possible when it is widely known that omitting relevant variables biases the remaining estimates? The answer depends upon how we interpret these simple MRA estimates of publication selection bias. When heterogeneity is largely selected, simple MRAs may correctly filter this selected heterogeneity as well as selected random errors.

To explore this issue, we return to our simple MRA model (4.1) and assume the worst case – that some other factor, X, affects the reported estimates and is also correlated with the standard error (SE).[25] This gives

$$effect_i = \beta_0 + \beta_1 SE_i + \beta_2 X_i + \varepsilon_i, \qquad E(X_i) = \gamma_0 + \gamma_1 SE_i \tag{6.15}$$

When MRA model (4.1) is estimated without including X_i, $E(\hat{\beta}_1) = \beta_1 + \gamma_1\beta_2$. One interpretation of this second term, $\gamma_1\beta_2$, is omitted-variable bias; another is the portion of publication bias operating more indirectly through the variable X. β_1 may be seen as the "direct" publication bias that results from resampling and re-estimation when the researcher's first estimate proves insignificant or of the "wrong sign." However, more typical in econometrics, researchers will respecify their econometric models by using a different set of independent variables, a different functional form, some new econometric technique, etc. Variations in such research dimensions create heterogeneity and are represented by X. When a study is imprecise (i.e. has a high SE), greater effort will likely be needed to obtain statistical significance. In these cases, a researcher is more likely to use some highly influential research dimension, X. Such selected heterogeneity contributes to publication selection bias and is evidenced by a correlation between X and SE. Thus, $\gamma_1\beta_2$ may be regarded as a component of publication bias.

By this interpretation, $\hat{\beta}_1$ from the simple MRA is not biased, but rather it estimates total publication bias coming from a variety of channels – recall $E(\hat{\beta}_1) = \beta_1 + \gamma_1\beta_2$. It is our view that this interpretation will often be appropriately true in economics and business research. Of course, this is only one interpretation. Another is that X imparts an important effect on the target phenomenon and any correlation between X and SE is just coincidence. In practice, it is impossible to know which interpretation is correct because it will depend on whether heterogeneity is created largely as a byproduct of publication selection or not. In our past experience over dozens of areas of research, simple MRAs of publication selection have always provided a satisfactory characterization of a given area of economics research, because these characterizations are also confirmed by more complex multiple MRA and methods.

As we show in Chapter 5, multiple MRA provides a more complex and nuanced estimate of publication bias. Nonetheless, for both of the multiple MRA examples reported in Chapter 5, the simple models of publication selection provide summaries that are corroborated by complex and robust multiple MRAs results. In the case of the value of a statistical life (VSL), both simple and complex methods find strong evidence of publication selection for statistically positive VSLs. Also, the associated estimates of corrected VSL are quite close to one another. Likewise for minimum-wage research. Both simple and complex multiple MRAs find evidence of selection for an adverse employment effect but no evidence of any practical employment effect, once allowance is made for this selection (see Chapter 5). Regardless of the interpretation chosen, we will still need to conduct several multiple meta-regression analyses to ensure that any simple interpretation is robust and/or to investigate how sensitive it is to more complex potential influences. Although substantial publication selection may allow one to ignore heterogeneity, in practice it is always a good idea to explore fully the more complex and nuanced multivariate landscape for the sake of robustness.

6.4.2 Nothing more than the least

We would also like to use a few words to defend the validity and rigor of the simple least-squares approach to meta-regression. In our view, there is little reason to use anything more sophisticated. The most sophisticated and rigorous statistical MRA model that we have found to be generally valid is the unbalanced "fixed-effects" panel MRA. Although these models begin with a complex error structure, they reduce to conventional linear regression and are efficiently estimated by OLS (recall Section 6.2). Either by adding dummy variables for studies or subtracting the study averages from all observed values, these multilevel panel methods reduce to OLS. Likewise for MRA models of publication selection, simple methods can be used effectively to filter out likely selection bias. In all MRA cases, we know there will be heteroskedasticity; thus, WLS should be the base MRA model. But then WLS is nothing more than OLS on weighted variables; recall MRA models (4.2) and (4.3). It is our view, confirmed by experience, that simple least squares is more resilient to the vagaries of research than more complex and seemingly more rigorous statistical methods.

However, the full arsenal of econometric techniques and methods can be fruitfully employed in meta-analytic applications. In the following chapter we explore the treatment of multiple effect sizes and the results from multiple meta-analyses. In the process, we offer a few examples of the rich opportunities for deeper understanding of research through the use of more sophisticated econometric approaches. In particular, we explore the estimation of the MRA using seemingly unrelated regressions (SUR) and thus three-stage least squares (3SLS).

Appendix: assumptions about error structures

In econometrics, it has long been proved that, as long as $\mathbf{\Omega}$ in (6.3) can be estimated up to some unknown proportion, σ^2, generalized least squares and weighted least squares estimates, as a special case, will have all of the desirable large-sample properties (e.g. Judge *et al*., 1982; Davidson and MacKinnon, 2004). That is, they will be unbiased and asymptotically normal with asymptotically unbiased standard errors. When $\mathbf{\Omega}$ is known, not estimated, up to this unknown proportion, the Gauss–Markov theorem proves that both the GLS and WLS estimators are best (minimum variance) linear unbiased estimates.

For example, Davidson and MacKinnon (2004: 261–2) are quite explicit about the issue of whether σ^2 needs to be constrained to be 1. That is, they show that if $\mathbf{\Omega}$ is replaced by $\sigma^2\mathbf{\Delta}$, where σ^2 is an unknown scalar, we still have all of the desirable GLS properties and that σ^2 can be estimated by the conventional OLS estimate of the variance of the regression errors (or MSE) for the transformed regression, equations (4.2) or (4.3) in our terms (Davidson and MacKinnon, 2004: 261). In practice, econometricians do not constrain σ^2 to be 1, because there is no need, nothing is gained by doing so, and the data themselves might give us a more realistic assessment of these variances. No doubt, it is for these reasons that statistical packages like STATA do not constrain MSE to be equal to 1 in their WLS routines.

Technically, there is a difference in the assumptions about the relation of between-study heterogeneity to within-study sampling variance between this general "fixed-effects" WLS-MRA and a "random-effects" MRA. With "random effects," between-study heterogeneity variance, τ^2, is assumed to be constant and independent of the sampling error, σ_i^2. In other words, the total variance (or unconditional sampling variance) is $\tau^2 + \sigma_i^2$. Our general WLS MRA assumes that the total variance is proportional to the conditional sampling error, σ_i^2, and thereby equal to $\sigma^2\sigma_i^2$. In the "random-effects" model, the variation among the weights, $1/\sigma_i^2$, is reduced by adding a constant value, τ^2, to σ_i^2, giving weights $1/(\tau^2 + \sigma_i^2)$.

With publication selection bias, we want the most precise estimates to be given a much larger weight, perhaps even more so than what $1/\sigma_i^2$ permits, to reduce publication selection bias – recall the *top 10* (Stanley *et al*., 2010). Thus, WLS will do a better job of giving the more precise effects a relatively larger weight than the "random effects" and thereby more fully compensating for publication bias. Furthermore, these two assumptions about how the total variance is or is not related to the conditional sampling variance are just that, assumptions of convenience. There is no reason, other than mathematical tractability, for assuming that between-study heterogeneity, τ^2, is independent of the sampling error, σ_i^2. With publication selection, between-study heterogeneity is likely to be dependent on σ_i^2; that is, less precise studies will, on average, engage in more model re-estimation and respecification to get the desired statistically significant results, and this might well affect the heterogeneity among the reported estimates. For all these reasons and others, we believe that "fixed-effects" WLS-MRA to be a viable benchmark specification and that the standard reported *SE*s from statistical packages such as

STATA and SPSS should be used.[26] That is, there is no need to divide the reported *SE*s by the square root of MSE.

No doubt, in the past, some WLS statistical packages have reported inappropriate statistics and that one should still confirm their validity before relying upon them. Nonetheless, recent versions of STATA and SPSS report correct WLS standard errors for the regression coefficients when σ^2 is not constrained to be 1. It is very easy to verify whether or not a given statistical package correctly reports WLS standard errors. Just have any statistical package compute equation (5.5) or (4.1) using WLS and $1/SE_i^2$ as the weights and then compare the standard errors of the regression coefficients to a simple OLS of (5.6) or (4.2), respectively. This comparison does not directly address Hedges and Olkin's (1985) issue about constraining σ^2 to be 1. That issue is best resolved by the research record itself.

Even with these modern improvements, WLS regression summary statistics should be considered suspect. For example, in WLS the reported R^2 refers to the standardized dependent variable (t-values) and not to the raw estimates. To get a R^2 in terms of the research estimates, recompute the residuals from the estimated regression coefficients and the raw data on $\mathbf{M}$ and $\mathbf{e}$. Then compare the variation in these residuals to the total variation in $\mathbf{e}$ (see any basic econometric text). Likewise, the square root of the MSE (or the standard error of the regression) may be reported in terms of these standardized values (SPSS) or in terms of the raw estimates (STATA). Thus, caution in interpreting statistical package results is always warranted.

7 Further topics in meta-regression analysis

In previous chapters we presented, derived and applied the basic MRA model and several variations. The aim of this chapter is to discuss a few additional dimensions of its structure and application. Section 7.1 explores some of the alternative applications of MRA in economics. In certain branches of economics, most notably environmental economics, MRA is used principally to derive improved estimates of key parameters, such as the willingness to pay. In other areas, the focus of MRA is mainly on the testing of competing economic theories, while other applications of MRA concentrate on modeling the heterogeneity among empirical findings. MRA is flexible enough to accommodate all of these facets of economics and business. In Section 7.2 we discuss the choice of MRA variables when there are more variables than observations. This is followed by a brief discussion of the functional form of the MRA in Section 7.3. We then discuss the use of MRA for identifying exclusion restrictions in Section 7.4. Section 7.5 looks at the forecasting performance of MRA in both time and space. Section 7.6 investigates the treatment of effect sizes that involve MRA models with interaction and non-linear terms.

The second part of the chapter explores the treatment of multiple effect sizes and the results from multiple meta-analyses. While most meta-analyses investigate a single effect size, there are many cases where researchers may be interested in the results of several related effects. In Section 7.7, we illustrate the use of systems estimators, such as seemingly unrelated regression (SUR) and three-stage least squares (3SLS), for dealing with multiple but related effect sizes. In Section 7.8, we show that the MRA model can be used to analyze the results from several meta-analyses of the *same* empirical phenomenon (the M2RA model). This section also discusses the results of meta-analyses of *unrelated* literatures.

7.1 Alternative applications of meta-regression analysis

In Chapter 4, we introduce a basic MRA, equation (4.1), that provides an estimate of the effect size corrected for publication bias:[1]

$$effect_i = \beta_0 + \beta_1 SE_i + \varepsilon_i \tag{4.1}$$

We regard this as the most basic MRA model. More informative is a general multivariate version of this basic MRA that enables conditional estimates of genuine effects, as well as publication and misspecification biases; recall equation (5.5):

$$effect_i = \beta_0 + \sum \beta_k Z_{ki} + \beta_1 SE_i + \sum \delta_j SE_i K_{ji} + \varepsilon_i \tag{5.5}$$

Versions of these MRA models can be used for a range of applications, such as summarizing and qualifying estimates of policy-relevant parameters, correcting these estimates for any number of potential biases inherent in observational economics research, testing economic theories, explaining heterogeneity, modeling the research process itself, and giving direction to future empirical investigation. These applications are not mutually exclusive. MRA can in fact be used to inform on all of these dimensions simultaneously. Indeed, we have argued throughout this book that MRA is best seen from a broad perspective encompassing several of these dimensions. In particular, we have argued that in order to derive improved estimates of policy-relevant parameters, it is essential that the MRA summarizes and explains past research, but also accommodates and minimizes publication and misspecification biases.

7.1.1 Improved parameter estimates

The focus of most meta-analyses is on deriving improved parameter estimates that are of direct use to policy makers. This is a major and important application of MRA. Examples of this prime directive include: the numerous meta-analyses on the value of a statistical life (VSL), environmental benefit transfer, and price and income elasticities of various commodities and taxes.

The large literature on VSL has spawned 14 meta-analyses and counting (e.g. Bellavance *et al.*, 2009). Most of these focus on estimating a single parameter, the value of a statistical life. Another important parameter in this literature is the income elasticity of VSL, which may be revealed as an ancillary MRA calculation. This elasticity is discussed further in Section 7.8.

Applications of meta-analysis in environmental economics often involve benefit transfer (e.g. Rosenberger and Loomis, 2000a; Shrestha and Loomis, 2001; Bateman and Jones, 2003; Brander *et al.*, 2006; Bergstrom and Taylor, 2006). Smith and Pattanayak (2002) ask whether this might not be environmental economics' "Noah's Ark." For benefit transfer, estimated coefficients from the MRA are used to predict the dollar value of sites that were not part of the original dataset. MRA can be used to predict valuations for "policy sites" using data on "study sites" and thereby saving much time and resources in conducting a new site-specific study (Shrestha and Loomis, 2001). But should they? And the likely errors in doing so are still an open question (Rosenberger and Stanley, 2006; Lindhjem and Navrud, 2008; Johnston and Rosenberger, 2010).

Many meta-analyses focus on elasticities derived from demand functions. Examples include own price elasticities for alcohol, tobacco, water, and energy.[2]

Precise estimates of such elasticities are very important for government taxation and health policies, and they can also be important for corporate decision making.

By averaging sampling errors and filtering publication and misspecification biases, the unconditional and conditional meta-estimates of effect sizes offer improved estimates of key parameters. The focus of the above three groups of meta-analyses has been predominately on parameter estimates. Thus, while most of the meta-analyses conducted in these literatures have investigated heterogeneity, the majority have largely abstracted from issues of publication bias. For example, of the 14 meta-analyses on VSL, only Day (1999) and Doucouliagos *et al*. (2012b) test for selection bias. We have seen in Chapters 4 and 5 that by ignoring publication selection bias, meta-analysis might result in faulty inference; in the case of both VSL and water price elasticities, controlling for publication bias greatly reduces the magnitude of the estimate. For benefit transfer in environmental valuation, ignoring publication selection can also cause serious bias, and correcting these biases usually makes the non-market values larger (Rosenberger and Stanley, 2006; Stanley and Rosenberger, 2009). However, ironically, using the FAT-PET-PEESE-MRAs that have been designed to accommodate and minimize publication bias and advocated in Chapter 4 can actually make the bias worse (Stanley and Rosenberger, 2009). As discussed in Chapter 4, this problem occurs when values are related to consumer surplus and derived from non-linear transformations of estimated demand coefficients. Simulations show that using a proxy for precision, the square root of the sample size, can go a long way towards reducing publication selection bias even in this perverse case.

Moreover, this type of meta-analysis has rarely tested economic theories. For example, none of the existing meta-analyses of VSL from wage-risk studies have explored the validity of the theory of compensating wage differentials. Similarly, the meta-studies on demand elasticities listed above do not test the validity of the law of demand. The focus of these meta-analyses has been on improving estimates of key parameters (e.g. the VSL and price or income elasticities), assuming that the underlying theories hold. We do not mean to suggest that these meta-studies are somehow fundamentally flawed; we wish merely to highlight alternative dimensions of MRA application.

7.1.2 Testing economic theories

Economic theory makes specific predictions about the distribution of empirical effects. Rival theories differ in terms of the direction, magnitude and the nature of the distribution of such effects. A natural application of the MRA is to test these rival theories. For example, neoclassical profit maximization in a competitive labor market predicts adverse employment effects arising from the minimum wage. Using meta-analysis, Card and Krueger (1995a) and Doucouliagos and Stanley (2009) test this prediction for the USA and find that the extant evidence does not support neoclassical theory. Like Card and Krueger (1995b), we speculate that perhaps alternative theories may offer a more accurate description of the data generating process, at least for the US teenage labor market. In a companion meta-analysis, Krassoi-Peach and Stanley (2009) find evidence in favor of the efficient wage hypothesis, and efficiency wages may be considered a "falsifying hypothesis" to the neoclassical competitive labor market theory (Popper, 1959).[3]

We believe that meta-analysis provides a viable platform from which to test economic theory rigorously and that only the comprehensive and objective perspective that meta-analysis offers can do so. For example, Stanley (2004, 2005b) reports linked meta-analyses which, together, constitute a sophisticated Popperian test of the natural rate hypothesis (NRH). Stanley (2005b) combines and meta-analyzes 34 tests of NRH and uncovers a clear pattern. Those tests that have more available information (larger degrees of freedom or sample sizes) find stronger evidence against NRH. This is exactly what statistical power would predict for a false hypothesis, and this interpretation is consistent with what the average of these tests of NRH indicates. The advantage of meta-analysis is that it integrates all the tests of a given hypothesis and can see across likely misspecification biases that might be present in any single econometric test.

However, not even the most comprehensive and rigorous meta-analysis, by itself, can provide a definitive or sophisticated "falsification" of an economic theory – at least not in a Popperian sense. Rather, a second "falsifying hypothesis" must be first confirmed:

> We shall take it as falsified only if we discover a *reproducible effect* which refutes the theory. In other words, we only accept the falsification if a low-level empirical hypothesis which describes such an effect is proposed and corroborated. This kind of hypothesis may be called a *falsifying hypothesis*.
>
> (Popper, 1959: 86–87)

Unemployment hysteresis is just this sort of "falsifying hypothesis" (Stanley, 2004). Unemployment hysteresis is the idea that the unemployment rate has a unit root or, in other words, is non-stationary. Shocks to the economy have very long-lived effects on unemployment. This hypothesis directly contradicts the NRH. If the unemployment rate is dominated by its own inertia, then there will be no "natural rate" of unemployment towards which unemployment gravitates.[4] Unemployment hysteresis is corroborated both by the observed rate of convergence of 99 persistence estimates from 24 studies and by the point towards which they converge. "Larger estimates of unemployment persistence are produced by models that use more information ($t = 9.03$; $p < 0.0001$) and are better specified" (Stanley 2004: 589). Thus, the NRH's falsifying hypothesis is corroborated by a second meta-analysis of a separate, but logically related, empirical literature.

We see great potential to use MRA to test rival economic theories and thereby to shape the development of economic theory. When the main interest lies in testing economic theory, the meta-analysis will likely focus on the value of a key parameter and the practical significance of this effect. This may also include testing the null of no relationship. However, in some cases, another value may be more economically relevant, such as whether an elasticity is 1 or a lagged unemployment coefficient is 1, and this too can be tested.

7.1.3 Meta-analysis to guide new estimates

MRA can be used to guide the development of econometric models. By definition, meta-analysis focuses on the analysis of the empirical studies reported by others. This function, however, is not limited to analyzing the past (i.e. what the existing literature has established). A good meta-analysis should serve as a guide for future empirical studies and even stimulate new meta-studies. Moreover, meta-analysis can be supplemented with original primary data analysis.

For example, Liu *et al*. (1997) provide original estimates of the VSL in Taiwan. They then proceed to offer a simple meta-analysis using data from developed countries. The aim of their meta-analysis is to compare their own econometric results to the results established by the broader literature. Doucouliagos and Ulubasoglu (2006) report a meta-analysis of the impact of economic freedom on economic growth. They then present their own primary econometric analysis which supported the conclusions from the meta-analysis.

One of the advantages of meta-analysis is that it applies a telescope to the empirical literature's findings and thus identifies gaps in empirical strategies and what is deemed to be best practice. Accordingly, a major benefit of meta-analysis is that it can open new directions in research. This work need not be left to other researchers. Indeed, it is our view that by dissecting an empirical literature, the meta-analyst is in an excellent position to undertake original, unique and informative primary data analysis. This is particular so for doctoral theses; empirical theses will benefit from including at least one chapter devoted to meta-analysis.[5]

7.1.4 Modeling the research process

In Chapter 5, we illustrate how MRA can be used to model the research process itself. Recall that the standard error terms (equation (5.5)) inform on how the publication selection process works in a given literature. Additionally, some of the **Z** vector variables quantify misspecification biases, another important aspect of the research process. However, MRA can be extended further. While it has rarely been used for this purpose, we see great potential in the use of the MRA for analyzing the historical evolution of economics research. By coding and subsequently analyzing an entire literature, the meta-analyst is able to address a range of issues, such as the choice of estimators and data used, who the leading researchers in the field are and how they have influenced others; and whether there is path dependence in the reported estimates. Stanley *et al*. (2008) provide a few examples, but the range of applications is truly enormous.

7.1.5 MRA: A multipurpose tool

As noted above, we do not see these alternative uses of MRA as mutually exclusive. There is nothing to prevent a well-structured MRA from testing rival economic theories, offering improved parameter estimates for policy, and modeling the process by which research in the field has been conducted.

7.2 Specification of the meta-regression analysis

A major issue in any econometric analysis is which variables to include in the econometric model. This is a major issue also for meta-analysis. Indeed, in some ways this can be more of a problem in meta-analysis than in primary econometric studies. Meta-analyses can quickly exhaust degrees of freedom. It is possible to end up with more study characteristics than actual observations. For example, the meta-analyst might want to control for differences in the measurement of the dependent and independent variables, the omission of relevant independent variables, differences in the composition of samples (country, firm and individual differences), differences over time, differences in functional form and estimator, as well as variables that can be constructed using information external to the empirical studies themselves.[6] Add to this variables that model the research process, and the number of explanatory variables expands rapidly. The meta-analyst might very well end up identifying, for example, 40 potential moderating variables but might be in possession of only 30 estimates.[7] Hence, it will often be necessary to omit some potential MRA variables. There are several ways to deal with this problem in practice.

Theoretically based exclusions are one way. Both theoretical and empirical literatures identify key issues. The MRA should, at least, attempt to explore these externally identified central issues. Hence, if it comes to the choice of variables that cover key issues versus other aspects that the meta-analysts might want to explore, preference should, in the first instance, be given to the former. The meta-analyst can, of course, always report alternative specifications. For example, the meta-analyst can report the results of an MRA that uses only the key variables identified by prior literature. When only a small empirical literature exists, we recommend this approach. This will enable the testing of key associations of interest and accommodating what prior research regards as important misspecification biases. Then the meta-analyst can report alternative MRA specifications for the sake of robustness and to explore several other effects especially if these were identified prior to calculating any statistics. Even though a more general model with all potential variables might be ruled out because of insufficient observations, it should still be possible for the MRA to answer a few of the key questions of a given area of research, while at the same time controlling for research dimensions found important in previous meta-analyses or in the research literature in question. Experience indicates that MRA reveals only a few important misspecification biases and research dimensions robustly even when there are ample degrees of freedom.

Another way is to choose factors reflected in the literature. Instead of coding every individual difference between studies, the meta-analyst can decide to test only those factors that are explored by a certain threshold number of studies. For example, if less than three studies use a certain control variable in the primary econometric analysis, then this might not be regarded as an important factor and can be ignored when degrees of freedom is a pressing issue.

Construction of new variables is a third way. In practice it is often possible to construct new variables that meaningfully capture a dimension of interest. For example, instead of including dummy variables for each country in a sample, regional dummy variables can be constructed (e.g. a variable for South America instead of separate dummy variables for Argentina, Chile, Brazil, and Peru). Similarly, a dummy variable can be constructed for systems estimators instead of separate dummy variables for, say, two-stage least squares, three-stage least squares, and so on. Krassoi-Peach and Stanley (2009) find that studies that make an effort to control for endogeneity of wages and worker productivity in any of several ways find much stronger efficiency-wage effects.

Another option is to use principal components analysis to collapse several variables into a newly constructed variable. This makes especially good research sense when the variables that are reduced into one all code for some similar research dimension such as the omission of important explanatory variables. Or sometimes it is possible to collapse multiple independent but related variables, say per-capita GDP and squared per-capita GDP, into one by subtracting a constant value from per-capita GDP before it is squared and only using this re-centered squared GDP per-capita variable. In this way, high multicollinearity can be avoided and yet the shape of the relationship can, nonetheless, be revealed.

Obviously, these strategies potentially increase the risk of omitted-variable bias in the MRA. However, when there are only a few estimates reported in a literature, the meta-analyst might have very little choice. Particular care should be taken with benefit transfer studies, where there is often a need to have relevant external information included in the MRA. Also, it is important to balance the needs for sufficient degrees of freedom with the need for an informative MRA. While it is important to consider issues explored by others, MRA can provide new insights into old questions. Hence, where there are sufficient estimates reported in a given area of research, the meta-analyst should strive to explore new research dimensions not considered in the current empirical literature.

However, in all cases, the meta-analyst should avoid data mining or the construction of variables that capitalize on chance or some quirk in the research literature. Thus, theory should be the meta-analyst's guide. As discussed in Chapter 6, the MRA is based largely on statistical theory. Anything that is known to shift the sampling distribution of the estimate in question (e.g. omitting a relevant variable in the original econometric study) should be included as an independent variable in the MRA. Unless there are many more reported estimates than such factors, this empirical literature may not be sufficiently mature to conduct a MRA.

7.3 Functional form of the meta-regression analysis

In addition to the specification of the MRA, the meta-analyst needs to consider the functional form of her model. Many MRAs will model effect sizes in levels, without any transformation of the variables. However, functional form might be important in some literatures, and researchers will need to consider the appropriate form this should take. The issue of functional form becomes particularly important when the effect size is measured as a dollar value. The most common transformations involve a log-transformation of the effect size or a log-transformation of one or more of the explanatory variables. For example, in the VSL literature, some estimates use the dollar value of VSL,[8] while some use the log-value. That is, some MRAs use a linear functional form, some use a double-log form, some use the lin-log form, while others use the log-lin form. The meta-analyst needs to decide whether to convert all estimates into a dollar figure or transform them into logarithms. When the commonly reported effect size is an elasticity, a semi-elasticity, or a partial correlation, such transformations are not normally necessary.

7.4 Exclusion restrictions

As already noted, when developing econometric models, researchers regularly face the difficult task of deciding which of the potentially large number of variables to include in their model. Often there is the need to balance the consequences of omitting a variable with pressure on degrees of freedom. This becomes even more pressing in the case of systems of equations, where potentially similar variables might influence a range of dependent variables. This raises the challenge of identification.

Meta-analysis can be of much assistance with identifying exclusion restrictions. By using the findings from existing meta-analyses, primary researchers might be able to exclude certain variables. That is, instead of resorting to theoretically based restrictions that lack empirical support, or worse still to *ad hoc* exclusion restrictions made for no other reason than necessity, meta-analysis can offer a more scientific and evidence-based approach.

For example, consider a primary researcher who wishes to estimate a system of equations that involves a growth equation and a human capital equation (among others). The researcher might be uncertain as to whether variables such as democracy, foreign aid and foreign direct investment (FDI) should be included in both equations, as theoretical models allow all these three variables to affect both growth and human capital formation. If it appears from existing meta-analyses that both democracy and foreign aid have no effect on growth, while FDI does, then the primary researcher can exclude the first two variables from the growth equation, and include them only in the human capital equation, which enhances the identifiability of the growth equation. That is, the findings from meta-analyses offer critical prior information that can be legitimately be used to shape primary econometric models.

7.5 Evaluating predictions from meta-regression analysis

All MRA models involve some inference and prediction, in terms of either time or space. For example, when MRA is used to test rival economic theories, the MRA findings explicitly apply for the time period studied. Researchers might also use the MRA coefficients to extrapolate forward in time. That is, the MRA coefficients can be used to predict the likely direction of the relationship under investigation, say for the next 5–10 years. Similarly, the MRA coefficients can be used to predict effect sizes in space. This is most commonly found in the benefit transfer of environmental values. That is, the MRA coefficients are established using data for certain sites/regions and then used to infer values for other sites/regions.

How successful/accurate are predictions from MRA? We can assess the performance of the MRA across both time and space in three ways: (1) How well does the MRA explain the research record? (2) Are the estimated MRA coefficients stable? (3) How well do the MRA coefficients transfer into related scenarios (e.g. benefit transfer)?

7.5.1 How well does meta-regression analysis explain the research record?

Like any regression, the explanatory power of the MRA is limited by the amount of variation in the underlying data that can be potentially explained – systematic heterogeneity. Because the dependent variable in an MRA is a statistical estimate, part of its variation from study to study is random sampling error and, hence, innately unexplainable.

The explanatory power of reported MRAs ranges from 0.08 to 0.98, depending on the research issue and the specification of the MRA.[9] Most of the MRAs do a reasonable job at explaining a significant portion of the heterogeneity in the research record. Indeed, half of the MRA models we have reviewed report an R^2 (or an adjusted R^2) greater than 0.50.[10] Unfortunately, few of the studies we have reviewed actually explore whether the remaining variation is solely due to random error. So it is difficult to assess fully how well extant MRAs explain the variation in reported economics and business research.[11] One exception is the study by Stanley (1998) whose meta-regression model explains all the heterogeneity, leaving only random sampling error unexplained.

7.5.2 Does meta-regression analysis withstand the test of time?

Describing the research record at a point in time is one thing, but how successful is MRA as a forecasting tool? That is, do the predictions of MRA models hold over time? One way of assessing this is to compare the predictions made by an earlier MRA with subsequent ones. Since MRA in economics is still relatively new, we only have a small number of examples of meta-analysis that have been reproduced.

One example comes from Doucouliagos and Paldam (2008), whose meta-analysis suggested that the effect of aid on growth was declining over time and was expected to continue to decline (see Figure 3.2). As a test of this, Doucouliagos and Paldam (2011a) updated their dataset and found indeed that the predictions of their earlier meta-analysis were correct; the effect of aid on growth continued to decline as predicted by the earlier meta-analysis.

A second example comes from the minimum-wage literature. Card and Kreuger (1995a) conducted the first meta-analysis of the employment effects of the minimum wage in the USA. They found that the evidence at that time pointed to no adverse employment effects. Doucouliagos and Stanley (2009) updated the Card and Krueger dataset and found that the earlier predictions held – the minimum wage in the USA has no adverse effect on employment.

In a third example, Stanley and Jarrell (1998) used a holdout sample of later studies on gender wage inequality to validate their MRA model. The models and findings over the two periods corresponded well, but the size of the prediction error was larger, as expected, in the holdout sample. Jarrell and Stanley (2004) updated and extended this rapidly expanding research on the gender wage gap and found largely consistent results, especially regarding their main findings. However, the affect and importance of a few moderator variables did change. Lastly, the central findings of Stanley and Jarrell (1998) of gender wage discrimination and its main findings were corroborated yet again by Weichselbaumer and Winter-Ebmer (2005) in their much larger international MRA of the gender wage gap, even though gender discrimination in different countries and cultures is likely to be quite different.

The results of MRA can change over time because the underlying relationships have changed over time and/or because new estimators and MRA modeling developments find something different. This means that it is entirely possible that earlier predictions are reversed, because new meta-analyses reveal new insights, correct past errors or omissions, or because new research reveals dynamic trends in the value of the genuine empirical effects. Some empirical literatures will be mature and well established, existing for relatively long periods of time, providing a rich research record. Others will be fairly dynamic, with new estimates emerging rapidly in ways that affect policy. Some literatures are growing exponentially. These differences in the pace and stability of reported estimates are a challenge for meta-analysis (or any informed review, systematic or otherwise). While the meta-data might be representative of research reality at the time the MRA was conducted, they need not be fully representative of the findings in the literature as new estimates roll out and as the phenomenon evolves. Many MRAs have found a significant time trend, confirming the dynamic nature of economics research. That is, parameter estimates can very well change over time because the underlying phenomenon may be dynamic or the methods used to study it may be evolving in important ways.

7.5.3 Do meta-regression analysis results transfer?

As already noted, applications of meta-analysis in environmental economics often involve benefit transfer functions (e.g. Bergstrom and Taylor, 2006; Shrestha and Loomis, 2001). That is, the estimated coefficients from the MRA are used to predict values of sites that were not part of the original meta-dataset. This is an application of MRA forecasting across site and space, rather than time. Rosenberger and Johnston (2009) discuss the various sources of error that might arise in the application of MRA for benefit transfer, and Shrestha and Loomis (2001) find that the average error of the MRA for benefit transfer to be around 24–30 percent. However, all approaches to benefit transfer can involve rather large errors, including meta-analysis (Rosenberger and Stanley, 2006; Lindhjem and Navrud, 2008); thus, there is still much to learn about how best to transfer the estimated benefit from one site to another.

7.6 Effects with interaction and non-linear terms

So far, we have considered only effect sizes associated with linear terms in econometric models.[12] Here, we consider the meta-analysis of empirical effects associated with interactions and non-linear terms, because comparable empirical effects are likely to be complex. As an example of these issues, consider the following econometric model:

$$Y_{it} = \alpha_0 + \alpha_1 H_{it} + \alpha_2 H_{it} \cdot K_{it} + \alpha_3 H_{it}^2 + \gamma \mathbf{X}_{it} + \varepsilon_{it} \tag{7.1}$$

where H is the key variable of interest for the meta-analysis and $\mathbf{X}$ is a vector of other factors that affect the dependent variable, Y. The interaction, $H_{it} \cdot K_{it}$, and non-linear terms, H_{it}^2, are important in identifying the *marginal* effect of H on Y, which in this case is given by: $\partial Y/\partial H = \alpha_1 + \alpha_2 K + 2\alpha_3 H$. The interaction term $H_{it} \cdot K_{it}$ measures the effect of H on Y conditional on the value of K. Similarly, the term H_{it}^2 causes the effect of H on Y to be conditional on its own value.

If this marginal effect can be calculated, it can be used in the meta-analysis.[13] The problem most meta-analysts will face is that this marginal effect is usually not reported. Typically, only the regression coefficients (α_1, α_2, α_3) and their standard errors will be available, rather than the marginal effect.[14] It appears that many authors are concerned only with the statistical significance of the individual interaction and non-linear terms rather than the practical significance of the overall effect. In some cases the marginal effect will be reported, and its standard error will not; it is rare for both to be reported. The biggest hurdle here is that the covariances are almost never reported.[15] This makes it difficult to include effects with interactions or polynomial terms and yet also accommodate publication selection. Meta-analysts could, however, use other weights, such as sample size or journal impact factors instead of standard errors. In practice, the meta-analyst will likely be forced to consider two strategies for dealing with this issue.

She can ignore the interaction and non-linear terms. The MRA can be applied to only those estimates from models that do not include any interactions or non-linear terms:

$$Y_{it} = \alpha_0 + \alpha_1 H_{it} + \gamma \mathbf{X}_{it} + \varepsilon_{it} \tag{7.2}$$

The main disadvantage here is that part (perhaps a very important part) of the literature is discarded, and this might introduce systematic bias into the MRA.

Alternatively she can conduct separate meta-analyses. The conditional terms can be used in a separate meta-analysis that explores the existence of a genuine empirical interactive term. For example, a meta-analysis can be carried out on the $H_{it} \cdot K_{it}$ term and separately for the H_{it}^2 term. If these meta-analyses reveal that these terms are important, then their results could be combined with those from the linear meta-analysis. If the meta-analyses indicate that these terms are not statistically significant, or if the size of the effect is very small and of little practical significance, then the interaction terms can be ignored and the MRA conducted only on the linear terms from all studies. That is, estimates of the effect of H on Y from equation (7.1) can be combined in the one meta-analysis, given that the interaction terms and the squared term can be taken to have no effect on Y. Note, however, that dummy variables should still be included in the MRA to identify those estimates derived from models with interactions and/or squares.

7.7 Multiple effect size analysis

Although it is not as serious as publication and misspecification bias, data dependence is another issue that should be explored in meta-regression analysis. In Chapters 4–6 we dealt with data dependence arising from multiple estimates of the same effect reported in the same study. At times, the researcher will be dealing with data dependence arising from correlations among multiple effect size measures. In such cases of multiple, yet related, outcome variables or effect sizes, the single equation MRA may no longer be appropriate, and we need to turn to fully multivariate MRA.

Table 7.1 summarizes the four scenarios that are likely to be encountered in practice. The standard scenario (A) arises when all studies included in the meta-dataset explore the same effect, such as the effect of the price of alcohol (X_1) on alcohol consumption (Y). The MRA can then be estimated using WLS, as discussed in Chapters 4 and 5. The overwhelming majority of extant MRAs use this type of data – they focus on a single effect size, such as the own price elasticity of alcohol.

Scenarios B, C and D involve multiple effect sizes arising from either multiple dependent variables (Y) or multiple explanatory variables (X). Scenario B reflects the case where the meta-analyst is interested in the effects of several explanatory variables. For example, the meta-analyst might want to conduct an MRA on the effects of the price of alcohol (X_1), income (X_2) and regulation (X_3) on alcohol consumption (Y). In this case, the meta-analysis involves three different effect sizes: the own price elasticity; the income elasticity; and the regulation elasticity. The dependent variable is the same in all cases (alcohol consumption) but the

Table 7.1 Structure of effect sizes

Effect size involves:		
Scenario	*Dependent variable (Y)*	*Explanatory variable (X)*
A	One	One
Multiple effect sizes		
B	One	Several
C	Several	One
D	Several	Several

effects differ because they involve different explanatory variables. It would not be appropriate to pool the different effect sizes (from X_1, X_2 and X_3) together and conduct a standard WLS-MRA on the pooled data because they measure conceptually different dimensions. Nonetheless, there is dependence in the effects of X_1, X_2 and X_3, because they are drawn from the same studies using the same data and dependent variable.

Scenario C occurs where multiple effect sizes arise from studies reporting the econometric analysis of more than one dependent variable, while scenario D involves analysis of related though different explanatory variables and different dependent variables. As an example of scenario C, consider the literature on economic freedom. Economic theory predicts that economic freedom will affect factor accumulation, especially investment in physical capital and human capital, as well as economic growth. Hence, theory predicts that economic freedom will impact on economic growth directly and also indirectly via capital accumulation. Empirical studies offer estimates on the effects of economic freedom on different dependent variables (growth (Y_1), human capital (Y_2), and physical capital (Y_3)). Researchers might be interested in analyzing all or some of these effects. In this case, there is a single explanatory variable (economic freedom) that has an effect on various dependent variables. Instead of running the MRA separately for each dependent variable, multivariate MRA can be applied allowing the joint estimation of several MRA equations, one for each dependent variable.[16]

Obviously, where there is interest in multiple effect sizes, the appropriate data will need to be collected and coded. In many cases, this information can be collected from the same pool of studies. In other cases, the information will need to be collected from different but overlapping literatures. Here we illustrate the use of two related estimators for dealing with such cases of cross-correlated MRAs, seemingly unrelated regressions (SUR) and three-stage least squares (3SLS).[17]

7.7.1 Seemingly unrelated regressions

In some applications, interest will lie in the analysis of the effects of different explanatory variables on different dependent variables. In many cases, these effects can be modeled as a sequence of individual MRAs. In some cases, however, the error terms in the set of MRA equations will be correlated. For example, there is a large literature on the price elasticity of alcohol consumption. Studies in this literature often report separate estimates for the price elasticity of beer, wine and spirits.[18] That is, they report more than one effect size – scenario D. If researchers are interested in the price elasticity of only one of these effects, say beer, then the single equation MRA is sufficient. However, efficiency gains are possible even when the meta-analysts wish to integrate only one effect, and interest often lies in more than one of these effect sizes. If we wish to estimate an MRA for each effect size, it might be statistically preferable to do so as a group rather than individually.

As a second example, consider the vast literature on the determinants of economic growth (scenario B). Prior meta-analyses have investigated the effects of individual variables, such as foreign aid, FDI, education, trade, institutions, democracy and inequality, one at a time and individually. Yet, we know that estimates of regression coefficients in a given regression relation will likely be correlated with the other estimated coefficients in this regression. This is what the variance–covariance matrix reflects.[19] However, as we discussed in Chapter 6, researchers rarely report the variance–covariance matrix in business and economics. Nonetheless, it would seem appropriate and probably more informative for meta-analyses to explore these multiple effects together, rather than in isolation. Doing so could, in effect, estimate the covariances of these estimated regression coefficients across the research literature using MRA.

Instead of providing estimates for each effect size separately, the MRA could be estimated as a set of seemingly unrelated regressions or even as a structural equation system. That is, instead of estimating a FAT-PET regression separately for each effect size, the MRAs can be estimated jointly. This joint estimation could also be conducted for the full multiple MRAs that include many *Z*- and *K*-variables. For the case of beer, wine and liquor, the SUR FAT-PET model would be:

$$\begin{aligned} BEER_i &= \alpha_0 + \alpha_1 SEbeer_i + \varepsilon_{1i} \\ WINE_i &= \beta_0 + \beta_1 SEwine_i + \varepsilon_{2i} \\ LIQUOR_i &= \gamma_0 + \gamma_1 SEliquor_i + \varepsilon_{3i} \end{aligned} \tag{7.3}$$

where *BEER*, *WINE* and *LIQUOR* are the own price elasticities for these different alcoholic beverages, *SEbeer*, *SEwine* and *SEliquor* are the standard errors for the elasticities of beer, wine and liquor, respectively, and ε_{ji} are the error terms for a given product and study, $i = 1,2,..., L$. In system (7.3), the right-hand-side variables are specific to each of the equations; every explanatory and every dependent variable is different. This corresponds to scenario D of Table 7.1. More generally, the MRA model might take the form of:

$$\begin{aligned} BEER_i &= \alpha_0 + \alpha_1 SEbeer_i + \alpha_2 \mathbf{Z}_i + u_{1i} \\ WINE_i &= \beta_0 + \beta_1 SEwine_i + \beta_2 \mathbf{Z}_i + u_{2i} \\ LIQUOR_i &= \gamma_0 + \gamma_1 SEliquor_i + \gamma_2 \mathbf{Z}_i + u_{3i} \end{aligned} \tag{7.4}$$

where the vector **Z** contains variables such as the country sample, the estimator used, the omission of important explanatory variables in the original study and the source of data. This system of equation can easily be extended to include the **K**-vector of variables interacted with standard error (recall Chapter 5). The **Z**-vector may contain the same type of variables across the three equations. Note that system (7.4) can easily be extended to include other equation-specific variables. While OLS still produces consistent estimates, the benefit of SUR over OLS is that it can result in efficiency gains;[20] SUR enables the cross-correlation in the errors to be incorporated in the estimation.[21]

We illustrate the application of SUR with 258 estimates on the own price elasticity of alcohol. The data come from 59 studies that report elasticities for all three types of alcohol.[22] Each of these studies reports estimates for all three types of alcohol. Within each study, estimates for the different types of alcohol are derived using the same time period, country and estimator. Hence, there is every reason to expect that there is significant data dependence within each study. Table 7.2 presents the results of the OLS estimates of each of the FAT-PET regressions (columns 1 and 2) and SUR (columns 3 and 4), for equation (7.3). While the coefficients are similar (with overlapping confidence intervals), the advantages of using SUR in this case are that it provides more efficient estimates and enables us to test whether the coefficients are the same across the equations. The Breusch–Pagan test strongly rejects the independence of the three equations ($\chi^2(3) = 128.63$; $p < 0.001$), suggesting that the SUR estimates are preferred to the OLS ones. The correlation between the errors in the beer and spirits equations is 0.64, between spirits and wine it is 0.26, and between beer and wine it is 0.13. That is, there is a rather strong correlation between the beer and spirits elasticities but not for the other combinations.

Testing the null hypothesis that all selection bias corrected alcohol elasticities are zero (H_0: $\alpha_0 = \beta_0 = \gamma_0 = 0$) is easily rejected ($\chi^2(3) = 1107.66$; $p < 0.001$), as is the null that the alcohol elasticities are identical (H_0: $\alpha_0 = \beta_0 = \gamma_0$, $\chi^2(3) = 485.43$; $p < 0.001$). Similarly, the null that the publication bias terms are zero (H_0: $\alpha_1 = \beta_1 = \gamma_1 = 0$) is also rejected ($\chi^2(3) = 26.99$; $p < 0.001$). Note that the MRA presented in Table 7.2 is meant as an illustration only. The full meta-regression analysis should obviously be extended to include a wide range of controls (versions of equation (7.4)) as discussed in earlier chapters. It is worth noting that this finding appears to run counter to Hunter and Schmidt (2004) who conjecture that publication bias is less likely where a literature revolves around multiple associations. This remains an important research area. Our own conjecture is that publication bias is a complex research phenomenon, and it is entirely possible that selection occurs across a wide range of associations. For example, it is plausible to suspect that a relatively “good” regression coefficient for beer might not be reported if the associated regression coefficient for wine is “undesirable”.

An example of the application of this estimator is Kotchen and Schulte’s (2009) meta-study of the fiscal impact of alternative land uses. The authors estimate a SUR model involving three equations for three land-use categories (residential, commercial and open space). The OLS results are not reported, but the authors note that OLS and SUR gave substantially similar results. A second example is the

Table 7.2 OLS versus SUR estimates of FAT-PET models

	Genuine effect (OLS) (1)	*Selection effect (OLS) (2)*	*Genuine effect (SUR) (3)*	*Selection effect (SUR) (4)*
Beer	−0.082 (−6.71)*	−2.226 (−8.63)*	−0.103 (−9.79)*	−1.95 (−8.00)*
Wine	−0.578 (−30.63)	0.674 (1.82)	−0.560 (−30.67)	0.523 (1.42)
Liquor	−0.136 (−9.54)	−2.520 (−7.85)	−0.168 (−14.07)	−2.181 (−7.04)

Notes: **t*-statistics in parentheses. Precision squared is used to weigh estimates. Sample size is 258.

study by Brons *et al*. (2008), who use SUR to meta-analyze the price elasticity of gasoline demand. They find that the SUR estimates are more plausible than those from OLS.[23]

The SUR estimator can also be used for data that match scenario B. For example, researchers might be interested in both the income elasticity of alcohol consumption and the price elasticity of alcohol consumption. Estimates of these two elasticities will come from the same studies, and involve the same dependent variable (alcohol consumption) but different explanatory variables (price and income). These two effect sizes could be meta-analyzed separately, by using SUR or through a structural system meta-regression equations.

7.7.2 Endogeneity in the meta-regression analysis

In some applications there will be two or more effect sizes that are endogenously determined. For example, consider the case of the literature exploring the effects of immigration on wages and the literature exploring the effects of immigration on employment. Longhi *et al*. (2010) note that in this literature there is an important issue of the joint impact of immigration on wages and employment. Empirical studies in this literature will report estimates of both the effects of immigration on wages and employment, and meta-analysts might be interested in both effects. In the case of the joint impact of immigration on wages and employment, Longhi *et al*. (2010) estimate the following MRA model:

$$\begin{aligned} {}^{w}b_{ij} &= \gamma_1 {}^{e}b_{ij} + \delta_1 M + \varepsilon_{1ij} \\ {}^{e}b_{ij} &= \gamma_2 {}^{w}b_{ij} + \delta_2 K + \varepsilon_{2ij} \end{aligned} \tag{7.5}$$

here ${}^{w}b_{ij}$ and ${}^{e}b_{ij}$ denote the elasticity of wages and employment with respect to immigration, respectively. In this framework, the elasticity of employment affects the elasticity of wages and vice versa. The authors accommodate this endogeneity by estimating the model using three-stage least squares and find that the results differ from OLS. Naturally, when estimating such systems it is important to be mindful of identification and to incorporate appropriate exclusion restrictions in the MRA.

This framework appears to be important when econometric studies focus on more than one effect size. This applies to both the estimates of the genuine empirical effect, as well as publication bias. For example, when estimating production functions, authors need to account for the marginal products of both capital and labor.[24] When a literature provides joint estimates of two or more related effects, the MRA should be modeled accordingly.

7.8 Meta-meta-analysis

As the number of meta-analyses grows, so does the pool of information on effect sizes that can potentially be compared within and between literatures. This wealth of information remains relatively untapped. Much can be learned from past

MRAs, and the tools of meta-analysis can be employed here also. We have seen in previous chapters and in literally hundreds of studies that MRA is an effective tool for modeling research heterogeneity. Perhaps, meta-meta-regression analysis can help identify patterns among meta-studies as well?

Surveys of meta-analyses can be accomplished by a traditional narrative review, by a meta-meta-analysis of the *same* effect sizes, or by a meta-meta-analysis across *different* effect sizes. Wherever possible, our own preference is for the more objective statistical analysis of a meta-analysis. It is important to note that we are not advocating the use of cumulative statistics for conceptually different hypotheses. Rather, we are advocating combining meta-analyses that test the same hypothesis (e.g. the effects of population on growth) or some other *comparable* effect between meta-analyses (such as the degree of publication bias).

7.8.1 Within-literature meta-meta-analyses

Multiple meta-analyses of the *same* literature will typically involve expansion of the research literature by adding newer studies, applying a different MRA estimator, or incorporating some other MRA methodology innovation. In short, they will tend to mirror the pattern observed in conventional applied econometrics.[25] As this body of knowledge grows, it becomes increasingly important to understand the differences among the reported meta-analyses findings, because there will be variation and seeming conflicts among them too. How should these differences be analyzed? What can we learn from the heterogeneity among meta-studies?

A few areas of economic research have received continued attention from meta-analysts. For example, the value of a statistical life has attracted 14 meta-analyses (Doucouliagos *et al*., 2012b) and there have been eight meta-analyses of wetlands valuations. An obvious way to put multiple meta-analyses into perspective is to offer a conventional narrative review of meta-analysis findings. Good narrative reviews can be quite insightful, and it is the only feasible alternative when there are only a small number of comparable meta-studies or estimates. However, as the number of meta-regression estimates on a given empirical magnitude grows, it will become increasingly difficult to understand their variation objectively or to map their multidimensional nature using qualitative reviews alone. This has been the clear lesson of empirical economics. Thus, meta-analysis of prior meta-analyses is the next logical step.

In a meta-meta-analysis, the unit of analysis becomes the MRA itself or one of its estimates. As in conventional meta-regression analysis, we will wish to identify those factors, other than random sampling error, that might explain differences among meta-results. In meta-meta-analysis, the dependent variable is either the principal finding in the individual MRA, or the coefficient on one of moderator variables used in the MRAs. For example, in the wetland meta-analyses, the main variable of interest is willingness to pay. While this might be the subject of the meta-meta-analysis, interest might shift to the effect that water quality has on the willingness to pay from one meta-analysis to another. Regardless of the chosen dependent variable, a set of moderator variables will need to be constructed. These might include the type of MRA estimator used, region, time period, correction for selection bias, and variations in the specifications of the MRA.

As with any meta-analysis, it is important to perform a comprehensive search for all prior meta-analyses to conduct a meta-meta-analysis. As an illustration of a meta-meta-analysis, consider Figure 7.1 which presents a funnel plot of 77 estimates of the income elasticity of the value of a statistical life (i.e. the percentage change in the value of a statistical life from a 1 percent increase in real income) from 13 meta-analyses.[26] This elasticity is important for cost–benefit analyses, as the value of a statistical life needs to be modified as income levels change over time or across regions. Figure 7.1 shows that the meta-studies report a fairly wide range of income elasticities. Most of these income elasticities fall between 0 and 1, but there are also a number of large elasticities, greater than 1.

A meta-meta-regression analysis model (M2RA) can be employed to identify the sources of heterogeneity:

$$\eta_{ij} = \beta_0 + \beta_1 \mathbf{X}_{ij} + \varepsilon_{ij} \tag{7.6}$$

where η_{ij} is the ith meta-estimate of the income elasticity from the jth meta-analysis, $\mathbf{X}_{ij}$ is a vector of explanatory variables and ε_{ij} is the error term. Table 7.3 presents estimates from this M2RA. For this illustrative example, moderator variables, the vector $\mathbf{X}_{ij}$, includes whether the meta-analysis corrects for selection bias, whether it considers only wage risk or only stated preference studies (with all types of studies combined used as the base) and whether the study was published.

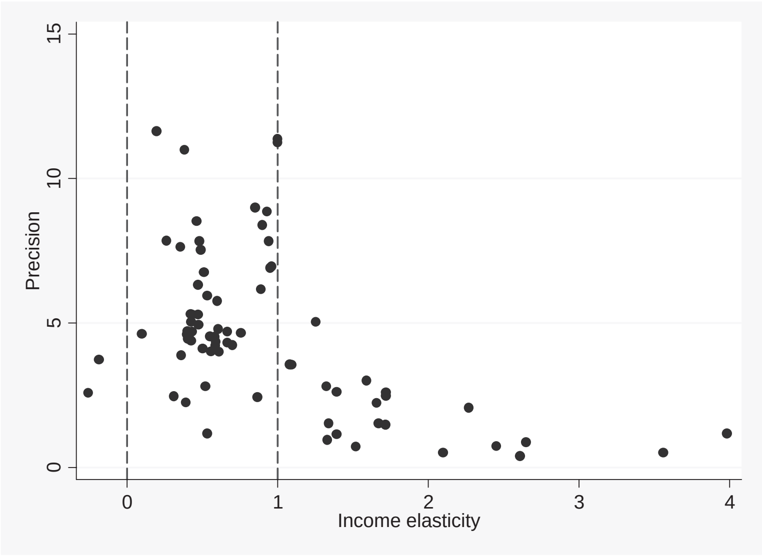


Figure 7.1 Funnel plot of meta-estimates of income elasticity of VSL

Table 7.3 WLS-M2RA of the income elasticity of VSL

Explanatory variable	*Mean (Standard deviation)*	*M2RA*
Constant		1.403 (5.34)*
Correction for selection bias	0.04 (0.19)	−0.705 (−2.63)
Wage-risk only	0.71 (0.46)	−0.402 (−10.63)
Stated preference only	0.05 (0.22)	−0.942 (−3.58)
Published	0.72 (0.45)	−0.482 (−1.81)
Number of observations		77
Number of meta-studies		13
Adjusted R^2		0.48

Notes: The dependent variable is the income elasticity of VSL estimates reported in meta-studies.
**t*-statistics are reported in parentheses using standard errors adjusted for data clustering.

Our aim here is simply to illustrate the possibilities of meta-meta-analysis, rather than to offer an exhaustive analysis of heterogeneity in this meta-literature. A more comprehensive analysis would consider other dimensions, such as alternative specifications of the MRAs used, differences in econometric methods and the country composition of the data.

Nonetheless, our illustrative M2RA model explains nearly half of the variation in the reported income elasticities of VSL. Meta-analyses that focus on compensating wage differential studies (wage risk only) and those that focus only on stated preference studies (stated preference only) find smaller income elasticities than meta-studies that include all approaches together. Meta-studies that are published also find smaller income elasticities, as do meta-studies that control for selection bias. The constant indicates that the income elasticity is 1.4 for unpublished studies that ignore selection bias and include all studies in the dataset. This suggests that life is a luxury good. However, this elasticity is halved if selection bias is considered and falls even further amongst published studies; therefore, life is not a luxury after all.

The right-hand tail of the funnel graph (Figure 7.1) suggests that finding a statistically positive income effect might be one of the dimensions that meta-analysts use to select the MRAs that they report. Hence, one extension of the M2RA presented in Table 7.3 would be to explore the degree of selection bias within meta-analyses. As we have said many times before, the full panoply of econometric methods and practices is available to meta-analysts, who are themselves econometricians.

7.8.2 Between-literature meta-meta-analyses

While it is meaningless to combine effect sizes of entirely different relationships, the information contained in meta-analyses of these relationships can be meaningfully compared for some applications. These associated surveys of meta-analyses can be conducted as a descriptive survey or narrative review or statistical meta-meta-analysis.

As an example of the survey approach, consider Rosenberger and Johnston (2009) who present two surveys of meta-analyses that draw upon the findings of meta-analyses of entirely different topics. First, they consider the direction of the trend coefficient in nine valuation meta-analyses. Their aim is to test whether there is "research priority selection sampling bias" (p. 414).[27] Second, they look at seven meta-analyses that have included a dummy variable for the publication status of the studies included in their samples. The aim of their second survey was to explore whether there are noticeable differences between the results of published and unpublished studies. In both cases, the surveys included meta-studies that dealt with conceptually different effect size. However, the focus of Rosenberger and Johnston's (2009) investigation is not the effect sizes themselves, but patterns among the effect sizes over time and between published and unpublished studies. That is, their focus is not on incomparable dependent variables but on comparable explanatory variables.

Meta-meta-analysis can be used to compare the size of the genuine empirical effect between related though distinct empirical literatures. A case in point is the enormous literature on the determinants of economic growth. Several meta-analyses have now investigated the effects of different variables on economic growth. One way to analyze the findings from these meta-studies is through SUR or 3SLS estimators, outlined in Section 7.7. However, this is possible only if the researcher has access to all the relevant data.

An alternative approach would be to compare the results in qualitative fashion. As an example of this, consider Table 7.4 which compares the findings of different studies. Some variables have been found to influence growth, while others have no effect at all. Unfortunately, the extant meta-studies use a wide range of measures of the size of the empirical effect. This makes it difficult to separate and compare meaningfully the relative practical and economic importance of the different effects on growth.

Meta-analysts often complain that many primary studies fail to provide sufficient information for subsequent users of their findings. One lesson we take from our

Table 7.4 Learning from meta-analyses, the determinants of economic growth

Author(s)	*Variable*	*Effect size used*	*Finding*
Doucouliagos and Ulubasoglu (2008)	Democracy	Partial correlation	No effect
Doucouliagos and Paldam (2008)	Development aid	Partial correlation	No effect
Abreu *et al*. (2005)	Convergence	Convergence rate	Positive effect
de Dominicis *et al*. (2008)	Inequality	Gini coefficient	Negative effect
Iamsiraroj (2009)	FDI	Partial correlation	Positive effect
Efendic *et al*. (2011)	Institutions	Partial correlation	Positive effect
Doucouliagos and Ulubasoglu (2006)	Economic freedom	Partial correlation	Positive effect
Mookerjee (2006)	Exports	*t*-statistic	Positive effect

meta-meta-analysis of the empirical growth literature (Table 7.4) is that there is also a problem within the practice of meta-analysts. Just as primary data researchers often do not consider the role of their individual study in the accumulation of knowledge and subsequent meta-analysis, so too meta-analysts have not really considered the role of their individual meta-study in the accumulation of meta-studies and subsequent meta-meta-analyses. But then, meta-analysis is still a relative recent and rapidly growing empirical approach; thus, its implications on general research practice have not yet been widely understood. In order to assist this process, we recommend that meta-analysts provide the MRA results on more than one effect size, wherever possible. For example, if partial correlations are used to ensure the largest sample size of comparable estimates, a meta-analysis of elasticities should also be reported even if it is for a reduced dataset. This provides subsequent meta-meta-analysis two effect sizes to add to others.

We see much potential for surveys of these sorts that communicate the findings of prior meta-analysis. However, in our view, M2RA can offer further insights about the underlying economic phenomenon (e.g. life is not a luxury), and be used to formulate new theories about the research process itself. Like meta-analysis, M2RA provides an empirical, evidence-based research framework to investigate research.

Meta-meta-analysis can also be applied to an examination of between literature factors. As an example of this type of M2RA, Doucouliagos and Stanley (2012) explore the links between the strength of the theoretical consensus regarding some economic phenomenon and the severity of publication bias observed among its reported estimates. Drawing upon 87 meta-analyses, covering nearly as many distinct empirical economics literatures and involving 3,599 studies and 19,528 effect sizes, we show that there is a robust link between the range of estimates that theory allows and what associated empirical studies report. The more contested theory is for a particular area of research, the less selection bias we find in this literature. Or from the opposite perspective, the stronger the theoretical consensus about a given economic effect, the larger is the observed publication bias. Ironically, the more strongly economists agree on a given phenomenon, the greater its empirical distortion. For example, we find that own price elasticities of demand tend to be highly inflated. Because such elasticities are central to many issues of government policy (e.g. energy, minimum wage and environmental policy), the practical implications of our M2RA are huge.

7.9 Summary

In this chapter we have explored several aspects, complications and potential applications of meta-regression analysis for economics and business. Like econometrics, the findings from MRA have many scientific and practical uses, and their models have great flexibility and potential in the modeling of empirical economic research. Because MRA is a relatively young discipline, much of its potential remains underutilized. Thus the future of meta-regression analysis in economics and business is bright.

This chapter has outlined alternative uses of MRA, strategies for specifying an MRA model, the success of MRA predictions, and how MRA may be generalized to explain and estimate multiple empirical effects and multiple MRAs. This discussion is only a sketch of the many facets of meta-regression analysis, and much more remains to be explored and further developed.

MRA models have the potential to contribute to the understanding of research well beyond that of single effect sizes in isolated research literatures. We show how MRA can be used to assist with the analysis of multiple effect sizes. It can also be employed to analyze the findings of meta-analyses. As more meta-analyses become available, it will be important to assess their findings to guide both new primary studies and new meta-analyses. This presents new opportunities and challenges for meta-analysts.

8 Summary and conclusions

Contemporary research on any important topic tends to be vast, and its findings are often disparate and widely dispersed. This "flood of numbers" threatens to drown economic knowledge and sensible policy action (Heckman, 2001). Yet, without some objective and comprehensive understanding of economics research, informed policy is impossible.[1] Unfortunately, conventional narrative reviews are not up to the task. It is all too easy for conventional reviewers to ignore findings that do not fit into their theories or ideology. Without objective standards, it is a hopeless task to assess what the genuine message of a given conventional review might actually be. Hence, experienced researchers often discount narrative reviews. And yet, there is an indisputable need for research synthesis.

Meta-analysis offers a critical and objective methodology that integrates conflicting research findings and can filter out some of the biases routinely found among reported research results. In hundreds of applications, meta-analysis has proven to be effective at cleansing the often murky and muddy waters of the ever-growing pool of econometric results. We believe that meta-analysis is indispensable for a clear understanding of actual economic phenomena.

In the preceding chapters, we have shown how meta-regression analysis can be used to:

- summarize an area of empirical inquiry;
- quantify estimates of policy-relevant parameters;
- explain much of the large systematic variation routinely found in business and economics research;
- correct reported research for evident misspecification and publication selection biases;
- test rival economic theories;
- address policy questions and evaluate policy interventions;
- model the research process itself;
- model prior meta-analyses results (the M2RA);
- point the way toward fruitful approaches for future research.

These applications are not mutually exclusive. We have attempted to show in this book that MRA can, in fact, be used to inform on all of these dimensions simultaneously. The applications of these MRA methods are truly boundless, ranging from deriving estimates of the willingness to pay for wetlands, the effectiveness of a given government or corporate policy, consumer behavior, and even meta-analyses themselves (recall the M2RA model).

Our approach to MRA may be summarized by the following five steps.

Step 1 – Selecting and coding of estimates

As discussed in Chapter 2, coding is the most time-consuming and, hence, costly step. It is, however, necessary to invest this time to produce an original meta-analysis or even to expand upon prior meta-analyses. In the process of identifying relevant research and coding them, we recommend that researchers be:

- *inclusive* in the research results collected;
- *comprehensive* in identifying and coding differences among research results;
- *objective* in defining clear criteria for study inclusion/exclusion and for coding variables;
- *insightful* and creative in identifying factors which might drive the reported research.

Researchers must avoid relying on their own priors or the research community's norms regarding factors such as journal quality and the authors' institutional affiliations when developing criteria for study inclusion. Applying such a lens to the inclusion of research results is likely to distort a meta-analysis in unknown ways. If these or other potential dimensions of study "quality" are deemed to be important, they should be coded and included explicitly in the meta-regression analysis itself. An inclusive and replicable approach to the collection and coding of a research record offers a scientific basis for the evaluation of research and, hence, provides an objective evidence base to policy and the understanding of economics and business.

Step 2 – Summarizing research

In Chapter 3, we illustrated the use of the funnel graph (a scatter plot of precision, $1/SE$, on empirical effect size) to reflect the distribution of reported research. Graphs and descriptive statistics of the coded empirical measures should be explored. We have found funnel graphs to be an especially useful way of viewing research. While descriptive statistics and graphs should be seen as strictly exploratory, such pictures can paint a thousand words and reflect complex phenomena in an efficient and concise manner. In our experience, funnel plots have been particularly useful in detecting coding errors, the possibility of publication selection bias and the existence of heterogeneity. Nevertheless, interpretation of graphs is inevitably subjective and no substitute for rigorous statistical analysis.

Descriptive statistics, such as weighted and unweighted averages, are widely reported in meta-analyses. We have argued, however, that any single, unfiltered summary statistic, such as fixed-effects and random-effects weighted averages, should be reported and interpreted with great caution if there is any suspicion of publication selection bias.

Step 3 – Accommodating publication selection bias

Publication selection bias has been detected in a wide range of empirical economics literatures (Doucouliagos and Stanley, 2012). In many cases, this bias is so substantial that it materially distorts statistical inference and any resulting understanding of research. Experience shows that all reviews, including meta-analyses, are vulnerable to the effects of selection. Hence, it is prudent for the meta-analyst to treat this phenomenon seriously. It would be a pity to incur the cost and effort of identifying, collecting and coding a large research literature without at least allowing for the presence of selection bias.

Chapter 4 presents simple yet effective MRA methods for accommodating and correcting an empirical literature for this bias. Our choice of MRA models of publication selection are the FAT-PET-MRA and PEESE-MRA. The FAT-PET-MRA is

$$effect_i = \beta_0 + \beta_1 SE_i + \varepsilon_i \tag{4.1}$$

This meta-regression model is estimated using weighted least squares with precision squared used as weights. The model contains tests for both publication bias (the funnel-asymmetry test, FAT; H_0: $\beta_1 = 0$), and the presence of a genuine effect beyond publication selection (the precision-effect test, PET; H_0: $\beta_0 = 0$).[2]

The precision-effect estimate with standard error (PEESE) is the estimated β_0 in:[3]

$$effect_i = \beta_0 + \beta_1 SE_i^2 + \varepsilon_i \tag{4.4}$$

The PEESE provides a better estimate of the actual empirical effect corrected for publication bias, when there is one.

Regardless of the outcome of the FAT, the PET should be used to test for the existence of a genuine effect beyond potential contamination from publication bias. If the meta-analysis does not permit a rejection of H_0: $\beta_0 = 0$, then this area of research fails to provide clear evidence of any genuine empirical effect. The PEESE should be used to estimate the magnitude of the empirical effect when the PET provides evidence that one exists.

This area of publication selection bias requires further research and will no doubt continue to evolve. Indeed, our own approach to meta-analysis has changed over the years. For example, we no longer advocate the use of the MST test (Chapter 4) even in conjunction with other publication selection MRA methods. However, a consensus seems to be emerging that FAT-PET-PEESE and similar meta-regression models provide the best correction for publication bias (Stanley, 2008; Moreno *et al*., 2009a, 2009b, Rücker *et al*., 2011).

It is also very important to evaluate the size of the corrected effect for practical economic or policy significance. Statistically significant effects that are practically small need to be given correspondingly little weight.

Step 4 – Modeling heterogeneity

Although simple MRA models of publication selection often provide an adequate overall estimate of empirical effect corrected for publication bias, we can never categorically rule out the possibility that some strong systematic research heterogeneity will overwhelm any single-value research summary, making it irrelevant and perhaps misleading. Furthermore, in economics and business, we frequently need to understand and explain the large variation found among of our research findings. Often, these conditional dependencies have the greatest relevance for theory and policy. For example, it is not enough to know that aggregate international aid has no practical effect on a country's economic growth (Doucouliagos and Paldam, 2008, 2011a). International development agencies need to identify which conditions or approaches tend to have successful outcomes and which tend to be counterproductive.

Chapters 5 and 6 offer multiple MRA models that have been widely successful in explaining research heterogeneity and, in the process, ensuring that any simple MRA finding is robust to more comprehensive analysis. The WLS-MRA model, equation (5.5),

$$effect_i = \beta_0 + \sum \beta_k Z_{ki} + \beta_1 SE_i + \sum \delta_j SE_i K_{ji} + \varepsilon_i \tag{5.5}$$

incorporates research dimensions that explain both the reported heterogeneity among results (Z-variables) and the propensity that a given finding will be reported and published (*K*-variables).

In order to ensure the robustness of the relevant explanatory research dimensions, a number of alternative MRA model specifications and methods need to be explored and reported. Within-study dependence can be accommodated through cluster-robust and fixed-effects panel or multilevel MRA. However, we caution against random-effects panel or multilevel MRA models, as they are likely to be invalid in applications of meta-analysis to economics due to correlation between the random study effects and the moderator variables.

Step 5 – Guiding research and policy

A well-constructed meta-analysis in economics should commence with a solid understanding of the underlying theoretical debates and issues. An objective methodology for evaluating the available body of evidence offers tremendous scope for resolving long-standing theoretical debates and, hence, for the evolution and development of economic and business theories. Moreover, because meta-analysis shines a light on the research process itself, it can also guide new and original primary econometric analysis. Indeed, meta-analytic practice has only scratched these surfaces.

Pitfalls to avoid

A key aim of this book is to guide those new to meta-analysis in economics and business. In our own applications, we have at one time or another encountered the full gamut of hurdles and challenges faced in the analysis of observational data. If the above five steps are followed, then meta-analysis will proceed more smoothly and the novice can avoid the most common pitfalls and errors. Common errors to be avoided include the failure to:

- carry out an adequate search of the literature, including unpublished studies;
- report adequately the search procedures employed, including reference to previous meta-analyses, surveys and reviews;
- collect and code standard errors and/or sample sizes;
- describe the data fully with summary statistics and graphs;
- understand the importance of the independence among estimates and the related problem of multiple estimates from the same study;
- weight effect sizes and/or to adjust for heteroskedasticity;
- recognize the possible importance of outliers;
- examine and correct for publication bias;
- employ panel methods and calculate cluster-robust standard errors when appropriate.

This list is adapted from Nelson and Kennedy's (2009) excellent review of meta-analyses in environmental economics. It is our sincere hope that the advice and discussions given in this book will help novice researchers to avoid these pitfalls and thereby contribute to the rigor and acceptance of MRA in business and economics.

Coda

Meta-regression analysis gives economists and business researchers the ability to summarize and evaluate their areas of empirical inquiry using the same tools and methods that produced the research in question. Our approach empowers empirical researchers themselves to evaluate their fields of specialization using their own methods, or similar ones, rather than relying on historians or philosophers of science to document and evaluate progress in economics and business.

Decades ago, logical positivists and naïve falsificationists sought to employ strictly logical criteria to assess a scientific research program and to indicate fruitful directions for its future progress (Popper, 1959; Lakatos, 1970). Since then, economic methodologists and philosophers of science have taken a more "naturalistic" turn; that is, a "turn away from a priori philosophy and towards a philosophical vision that is informed by contemporary scientific practice" (Hands, 2001: 129). This is precisely what MRA offers empirical economics. A well-conducted MRA may also serve as the basis for an internal philosophical appraisal of the scientific progress obtained in a given area of business or economics research.[4] MRA's potential for deeper philosophical reflection and evaluation has also been largely untapped.

Rather than attempting to develop a new logic of "induction" or "adduction" (Blaug, 1980) to justify econometric inferences or employing some *a priori* normative methodology of empirical inquiry, MRA takes econometrics as it is. Using econometric methods, MRA summarizes, corrects, tests and evaluates empirical econometric results. Economists cannot object that this approach is inappropriate or invalid, unless they also reject empirical econometrics itself.

Of course, econometrics and meta-econometrics (MRA) have their limitations and weaknesses. But a rejection of econometrics, *en masse*, would be quite uneconomical, a great waste of the massive research resources used to produce it. Furthermore, we believe that econometrics often provides the best empirical evidence that we have for many important questions of economic theory and policy. Readers of this book will realize that we are not naïve proselytizers of econometrics. Rather, we believe that econometric analysis often offers much less than is claimed. However, its modest findings usually add up to much more than nothing. Even when little remains in a given area of empirical inquiry after likely misspecification and selection biases are accommodated, this too is important to know.

Understandably, many applications in meta-analysis in economics have followed methodologies borrowed from other disciplines, most notably medicine and psychology. However, we have pointed out in this book that the meta-analysis of econometric studies requires a fresh meta-analytic approach. The aim of this book is to map out some of the key components of such methodological advance and differentiation. First, we have highlighted the need to treat observational data differently from experimental data. Second, selection bias is as much, if not more, a problem in economics as it is in other disciplines. Hence, it is imperative that meta-analyses examine their areas of research for selection bias, just as they should explore likely misspecification biases. Third, we caution researchers from applying widely used estimators, such as random-effects weighted averages and random-effects MRA models, to econometrics estimates.

Meta-analysis in economics is still relatively young. With youth comes hope and great promise. Just as econometric practice has evolved, so too will meta-analysis. The challenge for researchers is to continue to apply existing methods, to question them, and to develop new and improved approaches as needed. Much work lies ahead. We need to understand more about the various MRA models, especially the *Z*/*K* multiple MRA framework. Many more simulations along the lines of Stanley (2008), Koetse *et al.* (2010), Callot and Paldam (2011) and Stanley and Doucouiagos (2011) are needed to validate and test MRA methodology. The replication of prior meta-studies is also important. Do the conclusions from prior meta-studies survive when re-estimated with the ever-growing economics and business research? Perhaps new insights and lessons can be drawn from such replications. Because much progress has been made in recent years, current "best practice" specification of MRAs and the variables they should include requires further validation. It therefore remains important for researchers to explore the sensitivity of alternative MRA specifications. While new paths have been discovered, many require exploration and further development.

We see an evolving, yet bright, future for meta-econometrics. Its past successes and enormous potential, on so many levels, guarantee that meta-econometrics will continue to extend its reach and be further developed. May it shine its light on your area of research.

Notes

1 Introduction

1. Doucouliagos and Stanley (2012) conduct a meta-meta-analysis on the magnitude of publication selection bias (see Chapter 4 below) and the degree to which there is a competition of ideas about the prevailing economic theory in question. We find that debate and theory competition reduce the severity of this bias and its distorting effect on policy-relevant empirical magnitudes.
2. Experiments with human subjects cannot eliminate all potential threats to the validity of their inferences either; however, it is likely that a well-designed RCT will avoid some of the confounding influences that remain a concern to econometric analysis.
3. There are of course model specification tests, and they can help to reduce this ambiguity. However, they are infrequently employed, and their results will always leave some substantial uncertainty. These tests must, by necessity, make some assumptions about generic specification, and, even in the best of circumstances, there will remain the possibility of committing a type I or type II error.
4. Dozens of economic meta-analyses have corroborated the existence and importance of publication selection bias. See Chapter 4 for relevant references and an extended discussion of publication selection.
5. We use this example only as one clear illustration of how ideology often influences economics. It is not meant to be a comprehensive statement about the causes of the "Great Recession."
6. For example, UK's Department for International Development, in concert with other agencies, has funded several dozen systematic reviews and meta-analyses in 2010–11.
7. For two decades, the UK has been a world leader in attempting to better align chief executive pay to the performance of their companies and the interests of their shareholders. UK's regulations are generally called "comply or explain," because compliance is not technically mandatory, but failure to comply with government regulations requires the company to explain why, publicly (Arcot *et al*., 2010).
8. See Chapter 3 for an extended discussion of the weaknesses of vote counting. Vote counts can be greatly distorted by publication selection, which is known to be a serious problem in economics and business (Chapter 4). Furthermore, when statistical tests have low power, the probability that vote counts come to the *wrong* conclusion *increases* as research accumulates (Hedges and Olkin, 1985).
9. Statisticians have long argued that Glass's *g* is biased and inefficient, largely due to using a poor statistical measure of within group variation (Hedges and Olkin, 1985). Nonetheless, more sophisticated versions of effect size based on *g* remain widely in use.
10. Here, we use the term "multivariate" in its broadest, multidimensional sense, as multiple regression or similar statistical analyses. In statistics, multivariate analysis is often limited to situations where multiple dependent variables are being jointly modeled, or what econometricians would call simultaneous or structural equation systems. Meta-

regression analysis can be employed in both of these ways. Chapter 7 discusses the use of multiple MRA equation systems (e.g. three-stage least squares) to explain related empirical effects. In other chapters, "multivariate" is meant only as a generic, non-technical substitute for "multiple" MRA.

11 To give an accurate count of the number of meta-studies over time would require the type of comprehensive literature searching that characterizes meta-analysis. See the next chapter (Chapter 2) for guidance on how to conduct such an exhaustive search. Because our intention here is merely illustrative, we have made no effort to provide a comprehensive search of meta-analyses in economics. Nonetheless, we conjecture that such a meta-meta-analysis would reveal a similar pattern of growth.

12 The VSL is not a measure of the value of an actual life. Rather, it is the marginal willingness to pay for infinitesimal risk reductions for different people that are aggregated to a single *statistical* life.

2 Identifying and coding meta-analysis data

1 In time, meta-analyses might also shape economic theory. Card and Kruger's (1995a) meta-analysis on the minimum wage is one example where meta-analysis has influenced economic theory.

2 Banzhaf and Smith (2007: 1014) advocate that applied econometricians adopt meta-analysis as a way of providing a "statistical sensitivity analysis" to their own *individual* studies, especially when a large number of models are estimated.

3 It is not uncommon to omit accidentally one or two studies. However, such random omissions are not the result of systematic bias and, hence, will rarely have any practical effect on the inferences drawn from meta-analysis.

4 The choice of such a year should not, however, be arbitrary. Rather, it should be based on sound methodological grounds or underlying economic phenomena. For example, meta-analysts could choose to focus only on growth studies using post-Cold War data, or only those studies that use a newer and much improved dataset.

5 Indeed, primary studies should *always* be read carefully!

6 In our own research, we have on occasions found it necessary to physically check journals, volume by volume and issue by issue.

7 http://www.hendrix.edu/maer-network/default.aspx?id=15090.

8 It is not uncommon for the initial search to bring up hundreds, sometimes thousands, of studies, most of which must discarded because they do not contain relevant empirical estimates or tests.

9 We recommend that all excluded studies be referenced, either in the meta-study itself (e.g. Nelson, 2004) or as an online appendix (e.g. Doucouliagos and Stanley, 2009). This list enables readers to independently assess the comparability of the studies included in the meta-dataset. It also reduces the workload for future meta-analysts.

10 Ordered probit meta-models have also been used. For example, Waldorf and Byun (2005) use this model to explore the effects of age structure on fertility, while Koetse *et al*. (2009) apply it to the investment–uncertainty relationship.

11 This seems to be a practice that occurred in some fields in the past. It does not appear to be an issue in more recent studies.

12 Some meta-analyses include unpublished papers but fail to test for the existence of publication selection bias. Doucouliagos and Stanley (2012) show that, unfortunately, the inclusion of unpublished papers does not guarantee the absence of publication bias. Authors appear to select estimates in unpublished manuscripts in preparation for subsequent submission to journals.

13 In the disciplines of information technology and engineering, conference papers, especially refereed conference papers, are highly regarded. These are far less important in economics, where journal articles are treasured.

14 This can be tested by subsequent meta-analyses of the same literature.
15 It might be comforting also to referees, though in practice, none of our own meta-analyses were rejected by referees because they did not include enough studies or estimates.
16 For example, Doucouliagos and Paldam (2008) include all unpublished studies in their analysis of the development aid literature, including conference papers. Doucouliagos and Stanley (2009) include unpublished working papers and government reports in their analysis of minimum wages.
17 In a meta-regression analysis, this can be accommodated by including a binary variable controlling for whether the study was published or not (see Chapter 5).
18 It is advisable to also record the page number or table number from which the estimate was derived: this makes it much easier if it is necessary to re-examine the raw data at a later date.
19 The time period studied and the country examined are both particularly important for benefit transfer (see Chapter 7).
20 For example, Gallet (2007) includes dummies that explore the difference between myopic and rational addiction models of alcohol consumption.
21 An early example of this was Smith and Kaoru's (1990b) study of the price elasticity of demand for recreational sites. Other studies that use these variables include Stanley and Jarrell (1998), Görg and Strobl (2001), Waldorf and Byun (2005), Mookerjee (2006), Disdier and Head (2008), Klomp and de Haan (2010), Doucouliagos and Stanley (2009), and Bellavance *et al*. (2009).
22 See, for example, Disdier and Head (2008) and Klomp and de Haan (2010).
23 Some of this information is available from the studies themselves, for example in footnotes and acknowledgments. In some cases, it might become necessary to check the websites of authors and other sources.
24 There are exceptions to this. In some cases, interest might lie in simple correlations, but this is not common in economics.
25 Standardized regression coefficients are rare in economics research as their primary aim is better achieved by elasticities.
26 The use of the partial correlation in economics is advocated in Doucouliagos (1995) and Djankov and Murrell (2002).
27 When dealing with estimates from fixed effects panel data models, it is important to ensure that df is adjusted for the number of independent variables included (both time and cross-sections). This information is often poorly reported. Fortunately, the calculation of r is robust to uncertainty about df. For example, a t-statistic of 1.04 and degrees of freedom of 162 yield a partial correlation of 0.08. The partial correlation remains at 0.08 for all values of df from 149 to 191, so the results are robust to imprecise values of df.
28 In some cases, for the sake of brevity authors report the absolute value of the t-statistic, so it is essential to ensure that the correct value (positive or negative) is used. The partial correlation should have the same directional association as the underlying economic effect.
29 While it is technically a "correlation", in some applications it can be interpreted as causation. The researcher will need to take care to determine whether any effect size can be interpreted as causation.
30 The t-statistic may achieve the same objective (see Section 2.3.6). However, it must be modeled with added caution.
31 In our experience, most often than not, the very large partial correlations are estimated with low precision, typically emerging from studies with a small sample.
32 However, Hunter and Schmidt (2004) and Schulze (2004) caution that Fisher's z-transform replaces a slight downward bias in r with a slight upward bias in r. See Schulze (2004) for a discussion and comparison of other transformations of zero-order correlations.

33 The calculations are straightforward if the necessary data are available. See Gujarati (1995) for formulae for calculating elasticities from different functional forms.
34 Although statistically significant, such a small correlation (0.08) may be regarded as practically negligible (Cohen, 1988).
35 In this case, the meta-analysis essentially involves the collection of regression coefficients.
36 For details on the delta method, see Valentine (1979), Greene (1990) and Papke and Wooldridge (2005). The Fieller method is discussed in Valentine (1979) and Hirschberg *et al*. (2008).
37 An alternative approach is to use regression analysis to estimate the relationship between sample size and standard errors, for those estimates for which standard errors are available. The estimated regression coefficients can then be used to estimate the remaining standard errors. See Bellavance *et al*. (2009) for an application of this procedure.
38 While in reality there is a distribution of elasticities, most scholars are content to focus on the elasticity evaluated at the mean of the sample.
39 The survey response rate can be used for survey-based studies.
40 In most cases this involves dividing the short-run response by the estimate of the adjustment coefficient or 1 minus the coefficient on the lagged dependent variable.
41 See, for example, Sethuraman *et al*. (2011).
42 See, for example, Mookerjee (2006), Coric and Pugh (2010), and Klomp and de Haan (2010).
43 Unfortunately, we have seen several meta-analyses that have inappropriately used the *t*-statistic as a dependent variable. Hence, we urge scholars to exercise care with the specification of such models.
44 Common errors made by research assistants include confusing standard errors and *t*-statistics, wrong signs on *t*-statistics, incorrect sample size and/or degrees of freedom, and incorrect coding of control variables.
45 For example, Excel's TINV function can be used to convert *p*-values into *t*-statistics.
46 A weighted average of the standard error will also need to be calculated.
47 Meta-analysts are not exempt from selection bias. For example, the majority of the meta-analyses conducted on the value of a statistical life (VSL) have deliberately excluded negative values. The effect of such selection is to artificially inflate the VSL (see Doucouliagos *et al*., 2012b).
48 In contrast, estimates from the same author derived using different methods are not independent, when the data from which they are drawn are the same.
49 Perhaps the "leading journals" publish findings that are characterized by a winners' curse (Young *et al*., 2008). Large effects are reported which are subsequently found to be inflated. The curse in this case falls on the consumer of the reported effect sizes.
50 They also change frequently, so a decision needs to be made whether to use the impact factors at a point in time, or those that existed at the time the study was published.
51 This assumes that only estimates for which a measure of precision is available are included in the meta-dataset. Researchers might want a more comprehensive dataset, by exploring all estimates, even if precision is not available. While this would enable the calculation of descriptive statistics and summary measures (see Chapter 3), it would restrict correction for publication bias (see Chapter 4), although in some cases sample size might be used instead of precision.
52 Some studies report rather worrying evidence that "better journals" might be more selective and less precise (Waldorf and Byun, 2005; Young *et al*., 2008).
53 There is a fourth type of dependence arising from multiple effect sizes reported within studies. We discuss this type of dependence in Chapter 7.

3 Summarizing meta-analysis data

1 Some authors present time series graphs of the number of studies/estimates in the literature, showing their evolution over time (e.g. Nijkamp and Poot, 2004; Doucouliagos and Paldam, 2006).
2 The most common measure on the vertical axis is precision. However, some authors use sample size (Peloza and Steel, 2005; Vista and Rosenberger, 2009). Sterne and Egger (2001) discuss a range of alternative measures.
3 The funnels do not have to be centered at zero; they can be centered around any value.
4 Note that this is the reverse of what is shown on a funnel graph. The relationship between funnel graphs and conventional meta-regression models is discussed in Chapter 4.
5 See Mekasha and Tarp (2011) and Doucouliagos and Paldam (2011a) for contrasting views on these studies.
6 We have noticed a rather worrying pattern among the majority of meta-analyses that have explored the time dimension in their data; effect sizes in economics appear to be declining over time. This phenomenon warrants investigation.
7 While the inverse of the variance is technically the optimal weight, in practice variance has to be estimated. It was noted in Chapter 2 that the standard error of the partial correlation is a function of the size of this correlation. This might lead to bias in the meta-averages and it might also be an issue for the FAT-PET (see Chapter 4). Hence, Hunter and Schmidt (2004) and Schulze (2004) recommend the use of sample size as weights for correlation effect sizes. Researchers can always use both variance and sample size to explore the robustness of the results. In our experience, this makes little difference in the meta-analysis of econometric studies.
8 Such sensitivity analyses are often requested by referees.
9 Confidence intervals can also be constructed using the bootstrap (Adams *et al*., 1997).
10 It is important to note that these terms, fixed-effects estimator and random-effects estimator, as used in meta-analysis, are simple weighted averages and not the more sophisticated panel models used in econometrics. The latter models are also used routinely in multiple MRA to accommodate data dependencies (see Chapters 4–6).
11 We also show in Chapters 4 and 5 how there is much systematic heterogeneity and publication bias in the research on the employment effects of the minimum wage. Thus, the FEE should be treated as suspect for statistical reasons as well.
12 This is calculated using the average reported standard error as our estimate of σ. However, we show that there is much publication selection for an adverse employment effect in Chapter 4. After accommodating publication selection, it is not clear that any adverse employment effect remains; see Chapter 5 and Doucouliagos and Stanley (2009).
13 There are other tests with associated MRA model branches in the “tree” of research-driven meta-analysis (see Chapters 5).

4 Publication bias and its discontents

1 For details on these meta-analyses, see Doucouliagos and Laroche (2009) for unions and profits; Doucouliagos and Paldam (2011b) for aid allocations and democracy; Iamsiraroj (2009) for FDI and growth; and Shen *et al*. (2005) for hospital ownership and costs.
2 Liu *et al*. (1997), Day (1999), Miller (2000), Bowland and Beghin (2001), Dionne and Michaud (2002), Mrozek and Taylor (2002), de Blaeij *et al*. (2003), Viscusi and Aldy (2003), Kochi *et al*. (2006), Dekker *et al*. (2008), Kluve and Schaffner (2008), Bellavance *et al*. (2009), Lindhjem *et al*. (2010), and United States Environmental Protection Agency (2010).
3 We initially thought that this extreme skewness might be due to an exception to funnel symmetry found among non-market environmental values (Stanley and Rosenberger, 2009). See Box 4.3. However, this exception to funnel symmetry is caused by a non-

linear transformation of an estimated regression coefficient. In contrast, the VSL is calculated from a simple linear transformation of the estimated coefficient on the probability of death (Bellavance *et al*., 2009). Thus the shape of this funnel graph is dictated by the shape and selection of the estimated regression coefficients for the probability of death.

4 For details on these fields see Abreu *et al*. (2005) for beta-convergence, Rose and Stanley (2005) for common currency, Gallet and List (2003) for tobacco elasticity, and Gallet (2007) for alcohol elasticity.

5 Elsewhere, we have argued that Card and Krueger's (1995a) methods for identifying publication bias are flawed. Nonetheless, more rigorous methods and extensive meta-analysis confirm their conclusions (Doucouliagos and Stanley, 2009).

6 We trimmed a few (50, or 3.4 percent) of the most extreme elasticities to reveal the distribution of wage elasticities. Any estimated elasticity whose absolute value is greater that 1.1 is omitted from Figure 4.6. Omitting equal percentages from each side, if anything, accentuates the asymmetry of this graph.

7 Specification searches can also include data searches and variations in econometric techniques. In some cases, the sample size is deemed to be too small to produce the needed significant effects and hence researchers acquire more data. In other cases, the sample size is too large to produce the desired effects, and hence researchers find reasons to remove "outliers."

8 In previous papers, we reversed the use of β_1 and β_0, because we have always started with the weighted least-squares equation (4.2) as our baseline, where the intercept and slope are reversed from (4.1). Although such arbitrary notation choices do not matter, we hope our current use will be clearer for those unfamiliar with MRA and will not confuse those who are.

9 Equation (4.1) can be derived, approximately, from the expected value incidental truncation $effect_i = Z\beta + \sigma \cdot \lambda(c) + e_i$, where $\lambda(c)$ is the inverse Mills ratio and σ is the standard error (Greene, 1990; Wooldridge, 2002). Unfortunately, the usual Heckman method for sample selection is unavailable to us because the first-step probit requires that we observe both reported and unreported estimates. However, because the inverse Mills ratio will itself depend on the standard error, publication bias is likely to be a more complex function of the standard error. Simulations show that using the variance (i.e the square of the standard error) in the place of SE in equation (4.1) provides a better corrected estimator (Stanley and Doucouliagos, 2011); see PEESE, below. Section 6.3 provides a formal exposition of the mathematical derivation of this MRA model as an approximation.

10 This comes from the fact that WLS may be estimated by dividing the original regression model, $effect_i = \beta_0 + \beta_1 SE_i^2 + \varepsilon_i$, by SE_i. When SE_i^2 replaces SE_i in (4.1) and we divide by SE_i, no intercept remains.

11 Doucouliagos and Stanley (2012) discuss how the magnitude of $\hat{\beta}_1$ can serve as an indicator of the size of the publication selection bias. The larger is the absolute value of $\hat{\beta}_1$, the greater is the publication selection, *ceteris paribus*.

12 The great majority of reported minimum-wage elasticities concern teenage employment. It is widely acknowledged that adult employment is much less affected by changes in the minimum wage.

13 If non-robust standard errors are used, the precision coefficient (−0.009) is statistically significant, though still practically meaningless.

14 Actually, Doucouliagos *et al*. (2005) detect publication bias for positive union-productivity effects among US studies, and Stanley (2005a) finds publication bias in both directions (for statistical significance).

15 We did not select our examples for this reason. They were chosen to have broad variation in the shapes of their funnel graphs and to come from very different areas of economics research.

16 Although it is not important for the current discussion, we are assuming the all research studies are estimating a regression coefficient, α_1. We use α to denote the regression coefficients from the primary studies to distinguish them clearly from the MRA coefficients, β. Section 6.1 more formally discusses the relevant sampling distribution of the estimated empirical effect, $\hat{\alpha}_1$, and shows how statistical theory determines the basic structure of the MRA models.

17 Because the relationship between reported empirical effect and its standard error will be linear when there is no genuine effect (see Section 6.3), we need to caution readers to use MRA model (4.1) or (4.2) to test for the presence of a genuine effect and for publication selection. The PEESE-MRA model (4.3) will, therefore, be misspecified when testing for the presence of a genuine effect. It is also rather poor at identifying publication selection, because there can be strong correlation between the two independent variables in equation (4.3).

18 Stanley and Doucouliagos (2011) simulate a conditional estimator, the coefficient on $1/SE$ in equation (4.2) when the PET is not passed and the coefficient on $1/SE$ in equation (4.3) when the PET is passed. This combined corrected estimator is better than either alone.

19 Even the WLS versions of these models, equations (4.2) and (4.3), use simple OLS after first transforming the MRA model.

20 The modern view is to consider all unobserved study effects as random and to model them as "random," in the traditional sense, if they are independent of the independent variables, and as "fixed" otherwise (Wooldridge, 2002). Section 6.2 discusses panel models in greater technical detail. Because many economists and software packages still use this terminology of fixed and random panel effects, we use it here as well. Applied researchers will be forced to choose between fixed and random panel or multilevel methods by their statistical software.

21 We discuss panel MRA methods further in Chapter 6 and show how they can control for any unobserved quality difference between studies. In other disciplines these models are called "multilevel" or "hierarchical" linear models.

22 In STATA, such fixed-effects WLS panel estimates are obtained by **Xi:reg effect SE i.studyid [aweight=precision_sq]** where **studyid** is coded as a different integer for different studies. This STATA command automatically creates the necessary study dummy variables (**i.studyid**) to estimate a fixed-effect panel model. **aweight=precision_sq** weights the squared errors by $1/SE_i^2$.

23 Multiple estimates also enable a richer analysis of heterogeneity, and these panel models can control for unobserved study quality dimensions (see Section 6.2.2).

24 Strictly speaking, Hedges' maximum likelihood method does not require the likelihood of publication to be a monotonically non-decreasing function of the complement of an estimate's *p*-value. But it should. If publication selection is related to *p*-values, then smaller *p*-values must have an equal or greater chance of being accepted for publication. Any other pattern of *p*-values is something other than publication selection, likely omitted heterogeneity or misspecification bias. When there is statistically significant evidence that higher *p*-values are more likely to be reported, we interpret this as evidence that MLPSE is misspecified and its MRA estimates potentially biased.

25 We know this to be the case because there are always some insignificant results reported in all of the areas of economic research that we have investigated. Economists are sufficiently contentious that someone will dispute nearly any claim. Also, the journals have a preference for novel findings. Thus, if some empirical result has been well established, there will be an incentive to report counter-evidence.

26 Technically, it is more precise to relate statistical power to the degrees of freedom available to the statistical test rather than its sample size. We do not make this distinction here because the difference between degrees of freedom and sample size is practically insignificant in economic applications. That is, it will not matter which one the meta-analysis uses in practice.

27 In past papers, one of us, Stanley (2005a, 2008) recommended using MST as one way to differentiate genuine effect from publication bias. The superior statistical properties of the FAT-PET-MRA, revealed by simulations, have convinced us that MST adds nothing to meta-analysis other than potential ambiguity. Yet we still believe that its motivating idea, statistical power, provides a breakthrough for addressing publication bias scientifically; hence our reluctance. But then science advances when old theories and approaches are found to be error or inferior. We continue to explore new statistical methods with the hope of proving them to be superior to the FAT-PET-PEESE methods advocated here.

5 Explaining economics research

1 We use the term "multivariate" MRA in this chapter as a generic substitute for multiple MRA, which is of course a type of multiple regression. Multiple MRA uses several explanatory variables to explain a single common measure of empirical effect. In Chapter 7, we explicitly address systems of MRA equations that are jointly used to explain multiple dependent measures of effect, but do not refer to these as "multivariate."

2 Simulations also suggest that using the sum of squared errors from our simple FAT-PET-MRA when it is not forced through the origin gives an adequate test for the presence of excess heterogeneity in most cases.

3 To take one example, the random-effects PEESE-MRA model (4.4) may be implemented in STATA by: **metareg** *effect SE_sq*, **wsse**(*SE*). The last expression specifies the within-study standard deviation, and *SE_sq* is the square of the reported estimate's standard error (or variance).

4 In our context, these models are more accurately described as "mixed-effects" multilevel, because they contain both "fixed effects" in the form of explanatory variables and a random study component.

5 Technically, the conventional WLS-MRA models that are used in economics are not "fixed-effects" MRAs. "Fixed-effects" MRAs as discussed in the broader statistics and medical research literatures assume that there is no between-study heterogeneity and that SE_i^2 fully estimates the uncertainty of each individual estimated effect (Hedges and Olkin, 1985; Konstantopoulos and Hedges, 2004). In contrast, conventional econometric WLS allows the data to estimate a multiplicative between-observation (or study) heterogeneity; see Section 6.1 and the Appendix to Chapter 6 for an extended discussion of this issue. For this reason, we prefer to call these "WLS-MRAs" and not "fixed-effects" MRAs.

6 Of course, we all know the values of all the estimates for our own research but not the unreported findings of others. Even unpublished working papers and dissertations might contain only selected results.

7 It would be very informative to have applied econometricians keep research journals of all of their analyses and decisions about which estimates to report to allow the meta-analyst to identify the variables that belong in the *K*-vector.

8 The PEESE version of the multiple MRA models of publication selection and systematic heterogeneity are still a bit "green," and simulation studies are needed to validate their desirable statistical properties. Furthermore, the purpose of PEESE is to solve the "extrapolation problem" as $SE \rightarrow 0$ in order to estimate a single corrected effect (see Section 6.3). With a multivariate explanatory MRA, there is no single corrected effect to estimate, and other *K*-variables can "bend" the *SE* relation as $SE \rightarrow 0$. Our limited experience makes us wonder about the reliability of the estimated MRA coefficients in (5.7). We suspect that the added "multicollinearity" of (5.7) compared to (5.6) may cause individual coefficients to be less reliable, because all the added *SE* terms will be correlated (but not linearly) with the $1/SE$ terms. Thus far, only a few studies have used these complex multiple MRA models. More experience and research on these methods are needed to validate their use.

9 It is statistically equivalent whether meta-analysts code for the inclusion or the omission of a particular variable from the estimated empirical model. However, the interpretation of the intercept and other meta-regression coefficients can differ greatly depending on which definition of these moderator variables is employed.

10 Female researchers tend to report smaller gender wage gaps, male researchers larger ones. Stanley and Jarrell (1998) speculate that researchers attempt to be objective and scientific by leaning away from their own group associations.

11 Florax and Poot (2007) identify 125 meta-analyses in economics. We conducted our own EconLit search in November 2009, identifying 431 meta-analyses in economics. This is very likely to be an underestimate, because other search engines find many more. For example, Business Source Premier identifies four times more papers than EconLit using the same search terms and limiting the results to "economics."

12 Doucouliagos *et al*. (2012b) show that even when the funnel plot is adjusted for heterogeneity it will still display a large degree of asymmetry and publication bias.

13 In Chapter 4, we reported our WLS-MRAs in the form of *t*-values vs. precision, such as in equation (5.6). Although this difference may cause some confusion, it is more important to interpret the MRA coefficients correctly. Thus, we have chosen to use the MRA form where the dependent variable is in the same units of measurements as the reported empirical estimates, and the MRA coefficients on the *Z*-variables will have these same units. Regardless of which way the MRA is run, one must be careful to interpret the MRA coefficients correctly and in terms of the reported estimated effects. Thus, in this chapter, we use the effect form of the MRA to make the interpretation clear.

14 This bias is calculated by the estimated coefficient on *SE* times the average *SE* in this literature, while the R^2 comes from the simple FAT-PET-MRA reported in Table 4.1.

15 Of course, a rich panel dataset could allow conventional econometrics to capture time and income effects on VSL. However, our MRA estimates measure the influence of these important factors above and beyond what the current econometric research literature offers. Meta-analysis can also point to where future research is needed.

16 "Arnould and Nichols (1983) argue that recipients of compensation usually demand lower salaries for increased risk of death. Empirical evidence has shown that the existence of compensation implies big reductions in wage levels (Fortin and Lanoie, 2000). These authors claim that studies omitting this variable must necessarily obtain biased results" (Bellavance *et al*., 2009: 451).

17 The confidence interval was calculated at the mean value using a statistical package. Most statistical packages will use estimated regression coefficients to "predict" values of the dependent variable. Here, we used a "mean" rather than an "individual" prediction interval because we are estimating the VSL for the typical worker under these broad conditions, rather than what the next econometric study might estimate it to be. Because there is always much variation between studies, the individual prediction intervals are much larger.

18 The intercept is also very different, but this is required to compensate for the larger coefficient on *LnIncome*.

19 Here, too, we use the MRA form (equation (5.5)), where the dependent variable is the estimated empirical effect, minimum-wage elasticities in this case. It is our hope that by using different forms of the same MRA model in Chapters 4 and 5, the interpretation of the MRA coefficients will become clear.

20 We do not wish to claim that there is actually a positive effect on employment from raising the minimum wage. However, meta-regression analysis fails to find any practically significant adverse effect. We will return later to this issue of using a multivariate MRA to predict the "best practice" estimate of minimum wage's employment effect.

21 This is calculated from the conventional linear restrictions test using the WLS-MRA reported in column 1 of Table 5.6. However, a likelihood ratio test based on the more sophisticated multilevel MRAs gives the same assessment.

22 The average publication bias in terms of the minimum-wage elasticity is calculated from the estimated value of $(\beta_1 + \sum\delta_j K_j)SE$ or $\beta_1 SE$ using sample means.
23 This might seem to contradict what was said previously, but it does not. Before, we argued that it would be inappropriate to substitute *all* of the sample means for both the *K*- and *Z*-variables into the estimated meta-regression equation. Doing so would just give us back the reported sample mean elasticity, which, as we have seen, contains considerable publication bias. We purposely set $SE = 0$ to remove publication bias.
24 Most economists would argue that published papers are of higher quality than those that are not. We are unconvinced because we have seen all too often how papers that are selected to be published have greater publication bias. For the sake of robustness, we accept that the "best" research will be published.
25 Neoclassical theory predicts a negative employment response in the long run; thus, it is often argued that a negative adverse employment effect will only appear after a lag. However, our MRA shows clearly that any adverse effect is lessened (or positive employment effect strengthened) when lagged employment effects are measured-see *Lag* in Table 5.6. If we were to consider lagged effects as part of "best practice," all of the above corrected estimates would be positive.
26 Stanley (2001) recommends a third method to accommodate within-study dependence-MRA of the average study effects. This method is illustrated in Chapter 4, and it works quite well for mature areas that have a large number studies. However, practice in the field has evolved. Current consensus among meta-analysts is to use all reported estimates in an effort to maximize the information available for MRA.
27 The multivariate PEESE version also gives essentially the same results, with the exception that $Un{\cdot}SE^2$ is not statistically significant.
28 The Breusch–Pagan LM statistic is distributed as chi-squared with one degree of freedom.
29 In Section 5.1.2 we argued that meta-analysts should use neither "fixed" nor "random" effects MRAs, but rather WLS-MRAs when these models do not have an explicit panel structure. That advice still stands. In this section, we use the terms "fixed" and "random" in the same way that they are used by econometrics and statistical packages in the context of panel models.
30 This will help minimize selection bias *within* meta-analysis.
31 A given research dimension can be identified as both a *Z*-variable and as a *K*-variable.

6 Econometric theory and meta-regression analysis

1 Econometric theory is so widely known and universally relied upon that these necessary assumptions often go unreported in applied work.
2 We are using $\boldsymbol{\alpha}$ here and in previous chapters to represent the regression coefficients in the primary research literature because $\boldsymbol{\beta}$ are MRA coefficients.
3 Here, too, the meta-analyst has an advantage over conventional econometrics. When there are small-sample biases (e.g. in estimating an AR(1) coefficient), the sample size can be coded and included in a MRA (along with many other research dimensions) to track and thereby minimize this small-sample bias (Stanley, 2004).
4 Because reported empirical estimates come from different datasets with different sample sizes and other sources of heteroskedasticity, they are expected to have different variances and hence widths as we go up or down the funnel graphs. Recall that precision ($1/SE$) is the vertical dimension of the funnel graph. This heteroskedasticity is, in fact, measured by precision.
5 It is possible for the effect of one type of bias to depend on the presence of some other type of bias, but this more complex heterogeneity can be accomodated by including interaction terms in the MRA.

6 We are assuming that the empirical effect in question is $\hat{\alpha}_1$ from equation (6.1). In this matrix representation, we simplify notation. $effect_i$, is reduced to **e**.

7 For this interpretation to be accurate, even in theory, the moderator variables need to be defined in a manner that $M_j = 1$ represents the presence of some potential bias and $M_j = 0$ its absence.

8 Of course, OLS coefficients are unbiased even when there is known heteroskedasticity, and OLS gives the same WLS estimates when the MRA model is divided by σ_i, see Equation (6.4). If some correction for heteroskedasticity is not made, then we know that the MRA standard errors are likely to be wrong (biased), and statistical inference would not be reliable. Furthermore, when we have a multidimensional structure to our research data, then cluster-robust standard errors should also be computed for this baseline WLS-MRA.

9 Of course, a more complex GLS structure may also be employed, and within-study dependence needs to be addressed. In the next section, we explicitly discuss how one can model within-study dependence and more complex variance–covariance structures. Here, we use the baseline MRA model (WLS) to sharpen our focus on the theory of meta-regression analysis.

10 Becker and Wu (2007) offer a much more sophisticated multivariate GLS approach that jointly estimates all regression coefficients in the original regression equations using the full variance–covariance matrix. However, their approach, though theoretically more complete, is impractical in economic or business applications because it requires that the entire variance–covariance matrix be routinely reported in the research literature.

11 For example, STATA's basic regression procedure (**regress**) can be used but with $1/SE_i^2$ specified as the analytic weights. Or, **regress e M [aweight = Precision_sq]**, where **Precision_sq** is $1/SE_i^2$.

12 We do not recommend, however, that meta-analysts use heteroskedasticity-robust standard errors with the OLS estimation of MRA model (6.2). In addition to attempting to adjust for heteroskedasticity, weighting by precision also serves other important roles (e.g. a means to minimize publication selection and to reflect research quality).

13 Here, we are discussing simple meta-data structures. When multiple estimates are routinely reported by each study a panel model must be used, and a simple WLS-MRA that does not explicitly allow for unobserved study effects will not be adequate. We return to the subject of panel modeling of multidimensional meta-data in the next section.

14 To emphasize the importance of weighting by precision, Chapter 4 introduces panel MRAs with the t-value as the dependent variable. Here, we use the simpler MRA form, equation (4.6). Nonetheless, one should still use $1/SE_i^2$ as the analytic weights.

15 Note that the meaning of the fixed and random effects here differs from the conventional use of these terms in meta-analysis, where they denote weighted averages.

16 The meta-analyst who wishes to see all of the separate study effects, δ_i, can use the **Xi** command in STATA.

17 "Random-effects" unbalanced panel models are also called mixed-effects multilevel models, because there are both random, v_s, and "fixed effects," $\sum_{j=1} \beta_j M_{jis}$.

18 Only when there are very few estimates per study is there reason to prefer a "random-effects" approach. The fixed approach loses $K - 1$ degrees of freedom, whereas the random approach loses only one. However, only in extreme cases should efficiency concerns be allowed to trump bias and inconsistency.

19 Some researchers might be distracted by this discussion of "study quality." No doubt many will see precision, choice of econometric technique, etc. as indicators of "quality." No matter how you define study quality, it does not affect the findings of fixed-effects panel MRAs. First, remember that the study average values of *observable* quality indicators (e.g. precision, choice of econometric methods) are subtracted in fixed-effects panels; thus, these components of study quality will have no effect. In

contrast, observable differences in quality from estimate to estimate within studies can have an effect, and this is estimated explicitly by panel methods. Our concern in this section is with *unobserved study-level* components of quality and their potential to bias MRA estimates when fixed-effects panel methods are not used.

20 In theory, instrumental variables methods could also be used here. However, this alternative approach requires an instrumental variable (IV) that is correlated with the moderator variables, which are correlated with study quality and hence with the composite error, and yet the IV must be uncorrelated with study quality. In practice, finding a good IV that is not correlated with study quality would likely be difficult and controversial. Another approach, of course, is to find some proxy for study quality, code it, and include it explicitly as a moderator variable in the MRA. This approach is likely to suffer from an opposite problem-the correlation with study quality not being high enough. Fortunately, proxies for study quality are unnecessary when multiple estimates are routinely reported by each study.

21 Technically, publication selection involves incidental truncation rather than a censored dependent variable, as Heckman (1979) addresses. However, the Heckman regression is so widely known that putting a "Heckman" or "Heckit" label on models of incidental truncation has become the conventional terminology (Green, 1990: 744; Wooldridge, 2002: 564).

22 The exact expression in the conventional Heckman regression is a little more complex; however, its selection bias term contains both of these components, σ_i and $\lambda(c)$.

23 Moreno *et al.* (2009a) also investigate statistical methods specifically designed to use with the log-odds ratio. Log-odds ratios are different than conventional economic and business measures of effect in that their standard errors are correlated with their effects by construction, even when there is no publication selection. Because Moreno *et al.* (2009a) only simulate log-odds outcomes, some methods that are specifically designed to accommodate this relation of log odds to its standard were comparable to our FAT-PET-PESSE estimates. Moreno *et al.* (2009a, 2009b) call PEESE "Egger var." The PEESE model, however, was first articulated in Stanley and Doucouliagos (2007).

24 This mathematical result can be confirmed by a simple thought experiment. When there is no effect but only statistically significant results are reported, we would expect that the reported *t*-values to be just a little over 2, in which case the expected effect would be about twice its standard error.

25 We assume a correlation of *X* with *SE* here, because it is the worse-case scenario. When these are independent, there is no omitted-variable bias or threat to using the simple MRA model of publication bias.

26 When there are multiple estimates reported for each study, cluster-robust standard errors should be computed for this benchmark WLS-MRA.

7 Further topics in meta-regression analysis

1 There are many examples of published MRAs that do not control for selection bias, and some MRA studies do not use weights. Here, we assume that WLS is used with weights $1/SE_i^2$.

2 See Gallet and List (2003) on tobacco, Wagenaar *et al*. (2009) on alcohol, and Brons *et al*. (2008) on gasoline demand.

3 For slightly different applications of MRA to the minimum wage literature see Todorovic and Ma (2008) and Boockmann (2010).

4 Unless, this natural rate rapidly adjusts to the past unemployment rate. But even this will not rescue NRH, because doing so removes all of NRH's well-known and important implications (Stanley, 2002).

5 VU University Amsterdam has a graduate economics program that, in fact, requires such meta-analysis chapters; see http://www.feweb.vu.nl/en/departments-and-institutes/spatial-economics/master-point/index.asp.

6 This is particularly important to benefit transfer studies, where it is essential to include data on variables that are not part of the primary studies but which cover important contextual factors, such as income and population (Johnston and Rosenberger, 2010).

7 This is one reason why we prefer to use all comparable estimates reported in studies rather than a single estimate per study. The problem is likely to be worse in a new and emerging literature where relatively few studies are available. Jensen and Würz (2006) and Jensen (2010) outline an interesting testing procedure for models that have more variables than data points.

8 Note that dollar values need to be converted into a common base year, and when international comparisons are made these values need to be adjusted for purchasing power parity.

9 For example, in their meta-analysis of tax and expenditure limitations, Ballal and Rubenstein (2009) report R^2 ranging between 0.08 and 0.39. In contrast, in their meta-study of technology spillovers from FDI, Wooster and Diebel (2010) report goodness of fit ranging between 0.39 and 0.98.

10 In their survey of 140 meta-analyses, Nelson and Kennedy (2009) found that the median adjusted R^2 was 0.44.

11 See Lipsey and Wilson (2001) on this issue.

12 We have also discussed log-log and log-linear models which only use a single regression coefficient to estimate effect (see Chapter 2).

13 Typically, these marginal effects are evaluated using sample means. In those cases where such estimates are available, they can be meta-analyzed.

14 In some cases, a test for the joint statistical significance of the interaction terms is reported. Where enough such tests are reported, they could be used as the effect size.

15 It might be possible for meta-analysts to collect the original data and try to independently estimate these covariances. However, this is likely to be a very time consuming process with no guarantee that the original estimate or variance–covariance matrix can be replicated.

16 Doucouliagos and Ulubasoglu (2006) present meta-analyses of two effect sizes: the effect of economic freedom on growth and the effect of economic freedom on investment. However, the authors treat each effect as separate multiple MRA.

17 For alternative approaches adopted in the multivariate analysis of medical research, see van Howelingen *et al*. (2002) and Riley (2009).

18 Primary studies estimate these as either a separate equation for each type of alcohol or as a system of equations.

19 Becker and Wu (2007) offer a theoretical MRA model that uses the variance–covariance matrices in a multiple equation GLS system to jointly estimate several regression coefficients. But their approach requires that the variance–covariance matrices be routinely reported.

20 However, we are unlikely to achieve such efficiency gains, unless we have reason to believe that the errors terms are correlated, perhaps because the estimates are for the same time period or from the same study. In all cases, we are still assuming that the WLS version of all these MRAs will be employed.

21 Here, we consider the case where there are an equal number of observations for each equation. More generally, the meta-analyst may face an uneven number of observations for each equation. Schmidt (1977) and Baltagi *et al*. (1989) show that non-overlapping observations can be discarded. STATA enables the estimation of both balanced and unbalanced SUR equations.

22 We constructed our own dataset by updating the searches indentified by several prior meta-analyses. There is a much larger literature that reports estimates for only some of these types of alcohol (e.g. Wagenaar *et al*., 2009).

23 The interesting feature of the Brons *et al.* study is that they develop an approach that enables them to use estimates of six different elasticities with unequal number of observations. For example, they have only three observations of the price elasticity of mileage per car compared to 158 elasticities of total gasoline demand. Using a set of linear identities and the SUR estimates, the authors are able to derive a rich set of results.

24 Because several effects are jointly estimated, publication bias may be more complex also.

25 Empirical economics rewards innovation and usually shuns replication. Hence, it is unfortunate and unlikely that there will be many exact replications of meta-analyses. Jarrell and Stanley (2004) and Viscusi and Aldy (2003) are exceptions.

26 The meta-studies do not use a consistent specification. Some use log-log, some use log-lin, while others use lin-log. We converted all coefficients into comparable elasticities.

27 If more highly valued resources are evaluated first then subsequent studies will report smaller values over time.

8 Summary and conclusions

1 We are aware that the responsiveness of policy to empirical evidence is lower than economists would wish it to be. However, informed policy still requires a reliable evidence base.

2 Recall from Chapter 4 that the equivalent form is to divide by SE_i and estimate $t_i = \beta_1 + \beta_0(1/SE_i) + v_i$ (equation (4.2)).

3 Equation (4.4) is also estimated using WLS. Equivalently, the meta-analyst can estimate $t_i = \beta_1 SE_i + \beta_0(1/SE_i) + v_i$ (equation (4.3)).

4 There is much more written about the philosophy of science and the methodology of economics than about MRA; thus, any brief summary will by necessity be incomplete. We do not claim that our few comments provide a comprehensive or "systematic" review of these deep, complex, and dynamic literatures. We wish merely to make the claim that MRA can be seen as providing a "positive" philosophical evaluation of empirical economics that is consistent with a substantial slice of contemporary philosophy of science. In our view, Mayo (1996) provides a realistic and rigorous philosophical foundation for statistical inference, and this grounding could be easily extended to econometrics.

References

Abreu, M., de Groot, H.L.F. and Florax, R.G.M. (2005) A meta-analysis of beta-convergence: The legendary two-percent, *Journal of Economic Surveys*, 19: 389–420.

Adams, D.C., Gurevitch, J. and Rosenberg, M.S. (1997) Resampling tests for meta-analysis of ecological data, *Ecology*, 78: 1277–83.

Akerlof, G.A. (1982) Labor contracts as partial gift exchange, *Quarterly Journal of Economics*, 97: 543–69.

Albers, S., Mantrala, M.K. and Sridhar, S. (2010) Personal selling elasticities: A meta-analysis, *Journal of Marketing Research*, 47: 840–53.

Alston, J.M., Marra, M.C., Pardey, P.G. and Wyatt, T.J. (2000) Research returns redux: A meta-analysis of the returns to agricultural R&D, *Australian Journal of Agricultural and Resource Economics*, 44: 185–215.

Andrews, E.L. (2008) Greenspan concedes error on regulation, *New York Times*, October 23.

Arcot, S., Bruno, V. and Faure-Grimaud, A. (2010) Corporate governance in the UK: Is the comply or explain approach working?, *International Review of Law and Economics*, 30: 193–201.

Arnould, R.J. and Nichols, L.M. (1983) Wage-risk premiums and worker's compensation: a refinement of estimates of compensating wage differentials, *Journal of Political Economy* 91: 332–40.

Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999) A review of estimates of the schooling/earnings relationship, with tests for publication bias, *Labour Economics*, 6: 453–70.

Ayers, I. (2007) *Super Crunchers*, New York: Bantam.

Ballal, S. and Rubenstein, R. (2009) The effect of tax and expenditure limitations on public education resources: A meta-regression analysis, *Public Finance Review*, 37: 665–85.

Baltagi, B.H., Garvin, S. and Kerman, S. (1989) Further Monte Carlo evidence on seemingly unrelated regressions with unequal number of observations, *Annales d'Economie et de Statistique*, 14: 103–15.

Banzhaf, S.H. and Smith, K.V. (2007) Meta-analysis in model implementation: Choice sets and the valuation of air quality improvements, *Journal of Applied Econometrics*, 22: 1013–31.

Bateman, I.J. and Jones, A.P. (2003) Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in U.K. woodland recreation values, *Land Economics*, 79: 235–58.

Becker, B.J. and Wu, M-J. (2007) The synthesis of regression slopes in meta-analysis, *Statistical Science*, 22: 414–29.

Begg, C.B. and Berlin, J.A. (1988) Publication bias: A problem in interpreting medical data, *Journal of the Royal Statistical Society*, *Series A*, 151: 419–63.

Bellavance, F., Dionne, G. and Lebeau, M. (2009) The value of a statistical life: A meta-analysis with a mixed effects regression model, *Journal of Health Economics*, 28: 444–64.

Bergstrom, J.C. and Taylor, L.O. (2006) Using meta-analysis for benefits transfer: Theory and practice, *Ecological Economics*, 60: 351–60.

Berwick, D.M., Calkins, D.R., McCannon, C.J. and Hackbarth, A.D. (2006) The 100,000 lives campaign: Setting a goal and a deadline for improving health care quality, *Journal of the American Medical Association*, 295: 324–27.

Bijmolt, T.H.A., van Heerde, H.J and Pieters, R.G.M. (2005) New empirical generalizations on the determinants of price elasticity, *Journal of Marketing Research*, 42: 141–56.

Bland, J.M. (1988) Discussion of the paper by Begg and Berlin, *Journal of the Royal Statistical Society*, *Series A*, 151: 450.

Blaug, M. (1980) *The Methodology of Economics*, Cambridge: Cambridge University Press.

Boockmann, B. (2010) The combined employment effects of minimum wages and labor market regulation: A meta-analysis. IZA Discussion Paper No. 4983.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009) *Introduction to Meta-Analysis*, Chichester: Wiley.

Bowland, B.J. and Beghin, J.C. (2001) Robust estimates of value of a statistical life for developing economies, *Journal of Policy Modeling*, 23: 385–96.

Brander, L.M., Florax, R.J.G.M. and Vermaat, J.E. (2006) The empirics of wetland valuation: A comprehensive summary and a meta-analysis of the literature, *Environmental and Resource Economics*, 33: 223–50.

Brons, M., Nijkamp, P., Pels, E. and Rietveld, P. (2008) A meta-analysis of the price elasticity of gasoline demand: A SUR approach, *Energy Economics*, 30: 2105–22.

Brown, S.P. and Stayman, D.M. (1992) Antecedents and consequences of attitude toward the ad: A meta-analysis, *Journal of Consumer Research*, 19: 34–51.

Burkhauser, R., Couch, K.A. and Wittenburg, D.C. (2000) A reassessment of the new economics of the minimum wage literature with monthly data from the Current Population Survey, *Journal of Labor Economics*, 18: 653–80.

Callot, L. and Paldam, M. (2011) The problem of natural funnel asymmetries: a simulation analysis of meta-analysis in macroeconomics, *Research Synthesis Methods*, 2: 84–102.

Capelle-Blancard, G. and Couderc, N. (2007) How do shareholders respond to downsizing? A meta-analysis. Paper presented at the MAER Network Colloquium, September 27–30, 2007, Sønderborg, Denmark.

Card, D.E. and Krueger, A.B. (1995a) Time-series minimum-wage studies: A meta-analysis, *American Economic Review*, 85: 238–43.

Card, D.E. and Krueger, A.B. (1995b) *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton, NJ: Princeton University Press.

Card, D.E., Kluve, J. and Weber, A. (2010). Active labor market policy evaluations: A meta-analysis, *Economic Journal*, 120: F452–77.

Chalmers, T.C., Matta, R.J., Smith, H. and Kunzler, A.M. (1977) Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction, *New England Journal of Medicine*, 297: 1091–6.

Charemza, W.W. and Deadman, D.F. (1997) *New Directions in Econometric Practice*, 2nd edn, Cheltenham: Edward Elgar.

Cohen, J. (1988) *Statistical Power Analysis in the Behavioral Sciences*, 2nd edn, Hillsdale, NJ: Erlbaum.

Colegrave, A.D. and Giles, M.J. (2008) School cost functions: A meta-regression analysis, *Economics of Education Review*, 27: 688–96.

Connor, J.M. and Bolotova, Y. (2006) Cartel overcharges: Survey and meta-analysis, *International Journal of Industrial Organization*, 24: 1109–37.

Cooper, H.M. and Hedges, L.V. (eds.) (1994) *Handbook of Research Synthesis*, New York: Russell Sage.

Copas, J. (1999) What works? Selectivity models and meta-analysis, *Journal of the Royal Statistical Society*, *Series A*, 162: 95–109.

Coric, B. and Pugh, G. (2010) The effects of exchange rate variability on international trade: A meta-regression analysis, *Applied Economics*, 42: 2631–44.

Dalhuisen, J.M., Florax, R.J.G.M., de Groot, H.L.F. and Nijkamp, P. (2003) Price and income elasticities of residential water demand: A meta-analysis, *Land Economics*, 79: 292–308.

Davidson, J. (2000) *Econometric Theory*, Malden, MA: Wiley- Blackwell.

Davidson, R. and MacKinnon, J.G. (2004) *Econometric Theory and Methods*, Oxford: Oxford University Press.

Day, B.H. (1999) A meta-analysis of wage-risk estimates of the value of statistical life', Centre for social and economic research on the global environment. Working Paper.

De Blaeij, A., Florax, R.J.G.M., Rietveld, P. and Verhoef, E.T. (2003) The value of statistical life in road safety: A meta-analysis, *Accident Analysis and Prevention*, 35: 973–86.

De Dominicis, L., Florax, R.J.G.M. and De Groot, H.L.F. (2008) Meta-analysis of the relationship between income inequality and economic growth, *Scottish Journal of Political Economy*, 55: 654–82.

De Long, J.B. and Lang, K. (1992) Are all economic hypotheses false? *Journal of Political Economy*, 100: 1257–72.

De Mooij, R.A. and Ederveen, S. (2003) Taxation and foreign direct investment: A synthesis of empirical research, *International Tax and Public Finance*, 10: 673–93.

Dekker, T., Brouwer, R., Hofkes, M. and Moeltner, K. (2008) The effect of risk context on the value of a statistical life: A Bayesian meta-model. Institute for Environmental Studies, Working Paper W08/23.

Demsetz, H. (1974) Two systems of belief about monopoly. In H.J. Goldschmid, H.M. Mann and J.F. Weston (eds), *Industrial Concentration: The New Learning*, Boston: Little, Brown and Company, pp.164–84.

Dionne, G. and Michaud, P.C. (2002) Statistical analysis of value-of-life estimates using hedonic wage method. Working Paper 02–01. Ecole des Hautes Etudes Commerciales, Montreal.

Disdier, A-C. and Head, K. (2008) The puzzling persistence of the distance effect on bilateral trade, *Review of Economics and Statistics*, 90: 37–48.

Djankov S. and Murrell, P. (2002) Enterprise restructuring in transition: A quantitative survey, *Journal of Economic Literature*, 40: 736–92.

Doucouliagos, C. (H.) (1995) Worker participation and productivity in labor-managed and participatory capitalist firms: A meta-analysis, *Industrial and Labor Relations Review*, 49: 58–77.

Doucouliagos, C. (H.) and Laroche, P. (2003) What do unions do to productivity: A meta-analysis, *Industrial Relations*, 42: 650–91.

Doucouliagos, C. (H.) and Laroche, P. (2009) Unions and profits: A meta-analysis, *Industrial Relations*, 48: 146–84.

Doucouliagos, C. (H.) and Paldam, M. (2006) Aid effectiveness on accumulation: A meta study, *Kyklos*, 59: 227–54.

Doucouliagos, C. (H.) and Paldam, M. (2008) Aid effectiveness on growth: A meta study, *European Journal of Political Economy*, 24: 1–24.

Doucouliagos, C. (H.) and Paldam, M. (2009) The aid effectiveness literature: The sad results of 40 years of research, *Journal of Economic Surveys*, 23: 433–61.

Doucouliagos, C. (H.) and Paldam, M. (2010) Conditional aid effectiveness: A meta study, *Journal of International Development*, 22: 391–410.

Doucouliagos, C. (H.) and Paldam, M. (2011a) The robust result in meta-analysis of aid effectiveness: A response to Mekasha and Tarp. Aarhus University, Department of Economics Working Paper 2011–15.

Doucouliagos, C. (H.) and Paldam, M. (2011b) Does development aid reward good behavior? A meta-analysis of the effects of human rights and democracy. Paper presented at the 2011 European Public Choice Society Conference, April, Rennes, France.

Doucouliagos, C. (H.) and Stanley, T.D. (2008) A tale of two biases. Paper presented at the Nancy Workshop on Meta-Analysis in Economics and Business, 2008, University of Nancy.

Doucouliagos, C. (H.) and Stanley, T.D. (2009) Publication selection bias in minimum-wage research? A meta-regression analysis, *British Journal of Industrial Relations*, 47: 406–28.

Doucouliagos, C. (H.) and Stanley, T.D. (2012) Theory competition and selectivity: Are all economic facts greatly exaggerated? *Journal of Economic Surveys*, forthcoming. Also available as Deakin University, Economics Working Paper, 2008–06.

Doucouliagos, C. (H.) and Ulubasoglu, M. (2006) Economic freedom and economic growth: Does specification make a difference? *European Journal of Political Economy*, 22: 60–81.

Doucouliagos, C. (H.) and Ulubasoglu, M. (2008) ‘Democracy and economic growth: A meta-analysis, *American Journal of Political Science*, 52: 61–83.

Doucouliagos, C. (H.), Laroche, P. and Stanley, T.D. (2005) Publication bias in union-productivity research, *Relations Industrielles/Industrial Relations*, 60: 320–46.

Doucouliagos, C., Haman, J. and Stanley, T.D. (2012a) Pay for performance and corporate governance reform, *Industrial Relations*, forthcoming. Available as Deakin University, Economics Working Paper No. 2010–04.

Doucouliagos, C. (H.), Stanley, T.D. and Giles, M. (2012b) Are estimates of the value of a statistical life exaggerated? *Journal of Health Economics*, 31: 197–206.

Duval, S. and Tweedie, R. (2000) A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis, *Journal of the American Statistical Association*, 95: 89–98.

Efendic, A., Pugh, G. and Adnett, N. (2011) Institutions and economic performance: A meta-regression analysis, *European Journal of Political Economy*, 27: 586–599.

Égert, B. and Halpern, L. (2006) Equilibrium exchange rates in Central and Eastern Europe: A meta-regression analysis, *Journal of Banking and Finance*, 30: 1359–74.

Egger, M., Smith, G.D., Schneider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test, *British Medical Journal*, 315: 629–34.

Feige, E.L. (1975) The consequence of journal editorial policies and a suggestion for revision, *Journal of Political Economy*, 83: 1291–6.

Feld, L.P. and Heckemeyer, J.H. (2011) FDI and taxation: A meta-study, *Journal of Economic Surveys*, 25: 233–72.

Fidrmuc, J. and Korhonen, I. (2006) Meta-analysis of the business cycle correlation between the Euro area and the CEECs, *Journal of Comparative Economics*, 34: 518–37.

Fischer, C. and Morgenstern, R.D. (2006) Carbon abatement costs: Why the wide range of estimates, *Energy Journal*, 27: 73–86.

Fisher, R.A. (1932) *Statistical Methods for Research Workers*, 4th edn, London: Oliver and Boyd.

Florax, R.J.G.M. (2002) Accounting for dependence among study results in meta-analysis: methodology and applications to the valuation and use of natural resources. Series Research Memoranda 0005, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.

Florax, R. and Poot, J. (2007) Learning from the flood of numbers: Meta-analysis in economics. Paper presented at the Aarhus Colloquium of Meta-analysis in Economics, September 27–30, Sandbjerg Manor, Sønderborg, Denmark.

Fortin, B. and Lanoie, P. (2000) Effects of workers' compensation: A survey. In G. Dionne (ed.), *Handbook of Insurance*. Boston: Kluwer Academic Publishers.

Galbraith, R.F. (1988) A note on graphical presentation of estimated odds ratios from several clinical trials, *Statistics in Medicine*, 7: 889–94.

Gallet, C.A. (2007) The demand for alcohol: A meta-analysis of elasticities, *Australian Journal of Agricultural and Resource Economics*, 51: 121–35.

Gallet, C.A. (2010) The income elasticity of meat: A meta-analysis, *Australian Journal of Agricultural and Resource Economics*, 54: 477–490.

Gallet, C.A. and List, J.A. (2003) Cigarette demand: A meta-analysis of elasticities, *Health Economics*, 12: 821–35.

García-Quevedo, J. (2004) Do public subsidies complement business R&D? A meta-analysis of the econometric evidence, *Kyklos*, 57: 87–102.

Gerber, A.S. and Malhorta, N. (2008) Publication bias in empirical sociological research, *Sociological Methods and Research*, 25: 1–28.

Gerber, A.S., Green, D.P. and Nickerson, D. (2001) Testing for publication bias in political science, *Political Analysis*, 9: 385–92.

Glass, G.V. (1976) Primary, secondary, and meta-analysis of research, *Educational Researcher*, 5: 3–8.

Glass, G.V., McGaw, B. and Smith, M.L. (1981) *Meta-Analysis in Social Research*, Beverly Hills, CA: Sage.

Görg, H. and Strobl, E. (2001) Multinational companies and productivity spillovers: A meta-analysis, *Economic Journal*, 111: F723–39.

Greenberg, D.H., Michalopoulos, C. and Robins, P.K. (2003) A meta-analysis of government-sponsored training programs, *Industrial and Labor Relations Review*, 57: 31–53.

Greene, W.H. (1990) *Econometric Analysis*, New York: Macmillan.

Greenspan, A. (2007) *The Age of Turbulence: Adventures in a New World*, London: Penguin Press.

Griliches, Z. (1977) Estimating the returns to schooling: Some econometric problems, *Econometrica*, 45: 1–22.

Gujarati, D.N. (1995) *Basic Econometrics*, New York: McGraw-Hill.

Hands, D.W. (2001) *Reflection without Rules*, Cambridge: Cambridge University Press.

Harlow, L.L., Mulaik, S.A. and Steiger, J.H. (eds.) (1997). *What If There Were No Significance Tests?* Mahwah, NJ: Erlbaum.

Hausman, J.A. (1978) Specification tests in econometrics, *Econometrica*, 46: 1251–271.

Havránek, T. (2010) Rose effect and the Euro: Is the magic gone? *Review of World Economics*, 146: 241–61.

Heckman, J.J. (1979) Sample selection bias as a specification error, *Econometrica*, 47: 153–61.

Heckman, J.J. (2001) Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture, *Journal of Political Economy*, 109: 673–748.

Hedges, L.V. (1992) Modeling publication selection effects in meta-analysis, *Statistical Science*, 7: 246–55.

Hedges, L.V. and Olkin, I. (1985) *Statistical Methods for Meta-Analysis*, Orlando, FL: Academic Press.

Hedges, L.V. and Vevea, J.L. (1996) Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model, *Journal of Educational and Behavioral Statistics*, 21: 299–332.

Hennessy, P. (2011) Prime Ministers unite against Tory right, *The Telegraph*, June 4.

Higgins J.P.T. and Green, S. (eds) (2008) *Cochrane Handbook for Systematic Reviews of Interventions*, Chichester: Wiley.

Higgins J.P.T. and Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis, *Statistics in Medicine*, 21: 1539–58.

Hirschberg, J.G., Lye, J.N. and Slottje, D.J. (2008) Inferential methods for elasticity estimates, *Journal of Econometrics*, 147: 299–315.

Holmgren, J. (2007) Meta-analysis of public transport demand, *Transportation Research, Part A: Policy and Practice*, 41: 1021–35.

Hopewell, S., Loudon, K., Clarke, M.J., Oxman, A.D. and Dickersin, K. (2009) Publication bias in clinical trials due to statistical significance or direction of trial result, *Cochrane Review*, Issue 1. http://www.thecochranelibrary.com

Huang, J., van den Brink, H.M. and Groot, W. (2009) A meta-analysis of the effect of education on social capital, *Economics of Education Review*, 28: 454–64.

Hunt, M.M. (1997) *How Science Takes Stock: The story of meta-analysis*, New York: Russell Sage Foundation.

Hunter, J.E. and Schmidt, F.L. (2004) *Methods of Meta-Analysis: Correcting error and bias in research findings*, New York: Sage.

Iamsiraroj, S. (2009) *FDI and growth*. Unpublished PhD dissertation, Deakin University.

Iyengar, S. and Greenhouse, J.B. (1988) Selection models and the file drawer problem, *Statistical Science*, 3: 109–17.

Jacobsen, J.P. (1994) *The Economics of Gender*, Cambridge, MA: Blackwell.

Jarrell, S.B. and Stanley, T.D. (1990) A meta-analysis of the union-nonunion wage gap, *Industrial and Labor Relations Review*, 44: 54–67.

Jarrell, S.B. and Stanley, T.D. (2004) Declining bias and gender wage discrimination? A meta-regression analysis, *Journal of Human Resources*, 39: 828–38.

Jensen, P.S. (2010) Testing the null of a low dimensional growth model, *Empirical Economics*, 38: 193–215.

Jensen, P.S. and Würz, A.H. (2006) On determining the importance of a regressor with small and undersized samples. Aarhus Working Paper No. 2006–8.

Johnson, H.G. (1975) *On Economics and Society: Selected Essays*, Chicago: University of Chicago Press.

Johnston, N.L. and Kotz, S. (1970) *Distributions in Statistics: Continuous Univariate Distribution*, New York: Wiley.

Johnston, R.J. and Duke, J.M. (2009) Characterizing welfare patterns associated with study-invariant factors: spatial data supplemented meta-regression. Paper presented at the Oregon State University MAER Network Colloquium, October 2009, Corvallis.

Johnston, R.J. and Rosenberger, R.S. (2010) Methods, trends and controversies in contemporary benefit transfer, *Journal of Economic Surveys*, 24: 479–510.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T.C. (1982) *Introduction to the Theory and Practice of Econometrics*, New York: Wiley.

Klomp, J.G. and de Haan, J. (2010) Inflation and central bank independence: A meta-regression analysis, *Journal of Economic Surveys*, 24: 593–621.

Kluve, J. and Schaffner, S. (2008) The value of life in Europe: A meta-analysis, *Sozialer Fortschritt*, 10: 279–87.,

Knell, M. and Stix, H. (2005) The income elasticity of money demand: A meta-analysis of empirical results, *Journal of Economy Surveys* 19: 513–33.

Kochi, I., Hubbell, B. and Kramer, R. (2006) An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis, *Environmental and Resource Economics* 34, 385–406.

Koetse, M.J., de Groot, H.L.F. and Florax, R.J.G.M. (2009) A meta-analysis of the investment-uncertainty relationship, *Southern Economic Journal*, 76: 283–306.

Koetse, M.J., Florax, R.J.G.M. and de Groot, H.L.F. (2010) Consequences of effect size heterogeneity on meta-analysis: A Monte Carlo experiment, *Statistical Methods and Applications*, 19: 217–36.

Konstantopoulos, S. and Hedges, L.V. (2004) Meta-analysis. In D. Kaplan (ed.) *Quantitative Methodology for the Social Sciences*, Thousand Oaks, CA: Sage Publications, pp. 281–300.

Kotchen, M.J. and Schulte, S.L. (2009) A meta-analysis of cost of community service studies, *International Regional Science Review*, 32: 376–399.

Krakovsky, M. (2004) Register or perish, *Scientific American*, 291: 18–20.

Krassoi-Peach, E. and Stanley, T.D. (2009) Efficiency wages, productivity and simultaneity: A meta-regression analysis, *Journal of Labor Research*, 30: 262–8.

Krueger, A.B. (2003) Economic considerations and class size, *Economic Journal*, 113: F34–63.

Laird, N. and Mosteller, F. (1988) Discussion of the paper by Begg and Berlin, *Journal of the Royal Statistical Society*, *Series A*, 151: 456.

Lakatos, I. (1970) Falsification and the methodology of scientific research programmes. In I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.

Leamer, E.E. and Leonard, H.B. (1983) Reporting the fragility of regression estimates, *Review of Economics and Statistics*, 65: 306–17.

Leonhardt, D. (2007) Economist's life, scored with a jazz theme, *New York Times Book Review*, September 18.

Lewis, H.G. (1986) *Union Relative Wage Effects: A survey*, Chicago: University of Chicago Press.

Lewis, S. and Clarke, M. (2001) Forest plots: Trying to see the wood and the trees, *British Medical Journal*, 322: 1479–80.

Lindhjem, H. and Navrud, S. (2008) How reliable are meta-analyses for international benefit transfers? *Ecological Economics*, 66: 425–35.

Lindhjem, H., Navrud, S. and Braathen, N.A. (2010) Valuing Lives Saved From Environmental, Transport and Health Policies: A meta-analysis of stated preference studies. Environment Directorate, OECD, February.

Lipsey, M.W. and Wilson, D.B. (2001) *Practical Meta-Analysis*, Thousand Oaks, CA: Sage.

Liu, J-T., Hammitt, J.K. and Liu, J-L. (1997) Estimated hedonic wage function and value of life in a developing country, *Economics Letters*, 57: 353–8.

Longhi, S., Nijkamp, P. and Poot, J. (2010) Joint impacts of immigration on wages and employment: review and meta-analysis, *Journal of Geographical Systems*, 12: 355–87.

Loomis, J.B. and White, D.S. (1996) Economic benefits of rare and endangered species: Summary and meta-analysis, *Ecological Economics*, 18: 197–206.

Lovell, M.C. (1983) Data mining, *Review of Economics and Statistics*, 65: 1–12.

McCloskey, D.N. (1985) The loss function has been mislaid: The rhetoric of significance tests, *American Economic Review*, 75: 201–05.

McCloskey, D.N. (1995) The insignificance of statistical significance, *Scientific American*, 272: 32–3.

Mayo, D. (1996) *Error and the Growth of Empirical Knowledge*, Chicago: Chicago University Press.

Mekasha, T.J. and Tarp, F. (2011) Aid and growth: What meta-analysis reveals, UNU-WIDER Working Paper No. 2011/22.

Melo, P.C., Graham, D.J. and Noland, R.B. (2009) A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics*, 39: 332–42.

Miller, T.R. (2000) Variations between countries in values of statistical life, *Journal of Transport Economics and Policy*, 34: 169–88.

Mishel, L. and Rothstein, R. eds. (2002) *The Class Size Debate*, Washington: Economic Policy Institute.

Mookerjee, R. (2006) A meta-analysis of the export growth hypothesis, *Economics Letters* 91: 395–401.

Moreno, S.G., Sutton, A.J., Ades, A., Stanley, T.D., Abrams, K.R., Peters, J.L. and Cooper, N.J. (2009a) Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study, *BMC Medical Research Methodology*, 9:2, http://www.biomedcentral.com/1471–2288/9/2.

Moreno, S.G., Sutton, A.J., Turner E.H., Abrams, K.R., Cooper, N.J, Palmer, T.M. and Ades, A.E. (2009b) Novel methods to deal with publication biases: Secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications, *British Medical Journal*, 339: 494–98.

Mrozek, J.R. and Taylor, L.O. (2002) What determines the value of life? A meta-analysis, *Journal of Policy Analysis and Management*, 21: 253–70.

Mundlak, Y. (1978) On the pooling of time series and cross section data, *Econometrica*, 46: 69–85.

Nelson, J.P. (2004) Meta-analysis of airport noise and hedonic property values: Problems and prospects, *Journal of Transport Economics and Policy*, 38: 1–28.

Nelson, J.P. (2011) Alcohol marketing, adolescent drinking, and publication bias in longitudinal studies: A critical appraisal using meta-analysis, *Journal of Economic Surveys*, 25: 191–232.

Nelson, J.P. and Kennedy, P.E. (2009) The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment, *Environmental and Resource Economics*, 42: 345–77.

Nijkamp, P. and Poot, J. (2004) Meta-analysis of the effect of fiscal policies on long-run growth, *European Journal of Political Economy*, 20: 91–124.

Nijkamp, P. and Poot, J. (2005) The last word on the wage curve? A meta-analytic assessment, *Journal of Economic Surveys*, 19(3): 421–50.

Oaxaca, R. (1973) Male-female wage differentials in urban labor markets, *International Economic Review*, 14: 693–709.

Oosterbeek, H., Sloof, R. and van de Kuilen, G. (2004) Cultural differences in ultimatum game experiments: Evidence from a meta-analysis, *Experimental Economics*, 7: 171–88.

Papke, L.E. and Wooldridge, J.M. (2005) A computational trick for delta-method standard errors, *Economics Letters*, 86: 413–17.

Pavlides, M.G. and Perlman, M.D. (2009) How likely Is Simpson's paradox? *American Statistician*, 63: 226–33.

Pearson, K. (1904) Report on certain enteric fever inoculation statistics, *British Medical Journal*, 2: 1243–46.

Peloza, J. and Steel, P. (2005) The price elasticities of charitable contributions: A meta-analysis, *Journal of Public Policy and Marketing*, 24: 260–72.

Popper, K.R. (1959) *Logic of Scientific Discovery*, London: Hutchinson.

Popper, K.R. (1963) *Conjectures and Refutations*: *The Growth of Scientific Knowledge*, New York: Basic Books.

Poteete, A.R. and Ostrom, E. (2008) Fifteen years of empirical research on collective action in natural resource management: Struggling to build large-n databases based on qualitative research, *World Development*, 36: 176–95.

Public Broadcasting Service. (2009) The warning, *Frontline*, October 20.

Ridhwan, M.M., De Groot, H.L.F., Nijkamp, P. and Rietveld, P. (2010) The impact of monetary policy on economic activity: Evidence from a meta-analysis. Tinbergen Institute Discussion Paper 10–043/3.

Riley, R.D. (2009) Multivariate meta-analysis: The effect of ignoring within-study correlation, *Journal of the Royal Statistical Society, Series A*, 172: 789–811.

Rose, A.K. and Stanley, T.D. (2005) A meta-analysis of the effect of common currencies on international trade, *Journal of Economic Surveys*, 19: 347–65.

Rosenberger, R.S. and Johnston, R.J. (2009) Selection Effects in Meta-Analysis and Benefit Transfer: Avoiding unintended consequences, *Land Economics*, 85: 410–28.

Rosenberger, R.S. and Loomis, J.B. (2000a) Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database, *Water Resources Research*, 36: 1097–107.

Rosenberger, R.S. and Loomis, J.B. (2000b) Panel stratification in meta-analysis of economic studies: an investigation of its effects in the recreation valuation literature, *Journal of Agricultural and Applied Economics*, 32: 459–70.

Rosenberger R.S. and Stanley T.D. (2006) Measurement, generalization, and publication: Sources of error in benefit transfers and their management, *Ecological Economics*, 60: 372–78.

Rosenthal, R. (1979) The "file drawer problem" and tolerance for null results, *Psychological Bulletin*, 86: 638–41.

Rücker, G., Schwarzer, G., Carpenter, J.R., Binder, H. and Schumacher, M. (2011) Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis, *Biostatistics*, 12: 122–42.

Rusnák, M., Havránek, T. and Horváth, R. (2011) How to Solve the Price Puzzle? A Meta-Analysis. CERGE-EI Working Paper No. 446. Available at SSRN: http://ssrn.com/abstract=1942999

Sala-i-Martin, X.X. (1997) I just ran two million regressions, *American Economic Review*, 87: 178–83.

Sally, D. (1995) Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992, *Rationality and Society*, 7: 58–92.

Sandy, R. and Elliott, R.F. (1996) Unions and risk: their impact on the compensation for fatal risk, *Economica*, 63: 291–309.

Scargle, J.D. (2000) Publication bias: The "file drawer" problem in scientific inference, *Journal of Scientific Exploration*, 14: 91–106.

Schmidt, P. (1977) Estimation of seemingly unrelated regressions with unequal numbers of observations, *Journal of Econometrics*, 5: 365–77.

Schulze, R. (2004) *Meta-Analysis: A comparison of approaches*, Göttingen: Hogrefe and Huber.

Sethuraman, R, Tellis, G.J., and Briesch, R.A. (2011) How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities, *Journal of Marketing Research*, 48: 457–71.

Shen, Y-S., Eggelston, K., Lau, J. and Schmid, C. (2005) Hospital ownership and financial performance: A quantitative research review. NBER Working Paper No. 11662.

Shinew, K.J., Floyd, M.F. and Parry, D. (2004) Understanding the relationship between race and leisure activities and constraints: Exploring an alternative framework, *Leisure Sciences*, 26: 181–199.

Shrestha, R.K. and Loomis, J.B. (2001) Testing a meta-analysis model for benefit transfer in international outdoor recreation, *Ecological Economics*, 39: 67–83.

Sidik, K. and Jonkman, J.N. (2007) A comparison of heterogeneity variance estimators in combining results of studies, *Statistics in Medicine*, 26: 1964–81.

Simons, R.A. and Saginor, J.D. (2006) A meta-analysis of the effect of environmental contamination and positive amenities on residential real estate values, *Journal of Real Estate Research*, 28: 71–104.

Smith, M.L. and Glass, G.V. (1977) Meta-analysis of psychotherapy outcome studies, *American Psychologist*, 32: 752–60.

Smith, V.K. and Huang, J-C. (1995) Can markets value air quality? A meta-analysis of hedonic property value models, *Journal of Political Economy*, 103: 209–27.

Smith, V.K. and Kaoru, Y. (1990a) What have we learned since Hotelling's letter? A meta-analysis, *Economics Letters*, 32: 267–72.

Smith, V.K. and Kaoru, Y. (1990b) Signals or noise? Explaining the variation in recreation benefit estimates, *American Journal of Agricultural Economics*, 72: 419-33.

Smith V.K. and Pattanayak, S.K. (2002) Is meta-analysis a Noah's Ark for non-market valuation? *Environmental and Resource Economics*, 22: 271–96.

Stanley, T.D. (1998). New wine in old bottles: A meta-analysis of Ricardian equivalence, *Southern Economic Journal*, 64: 713–27.

Stanley, T.D. (2001) Wheat from chaff: Meta-analysis as quantitative literature review, *Journal of Economic Perspectives*, 15: 131–50.

Stanley, T.D. (2002) When all are NAIRU: Hysteresis and behavioral inertia, *Applied Economic Letters*, 9: 753–57.

Stanley, T.D. (2004) Does unemployment hysteresis falsify the natural rate hypothesis? A meta-regression analysis, *Journal of Economic Surveys*, 18: 589–612.

Stanley, T.D. (2005a) Beyond publication bias, *Journal of Economic Surveys*, 19: 309–45.

Stanley, T.D. (2005b) Integrating the empirical tests of the natural rate hypothesis: A meta-regression analysis, *Kyklos*, 58: 611–34.

Stanley, T.D. (2008) Meta-regression methods for detecting and estimating empirical effect in the presence of publication bias, *Oxford Bulletin of Economics and Statistics*, 70: 103–27.

Stanley, T.D. and Doucouliagos, C. (H.) (2007) Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. School Working Paper, Economics Series 2007–11, Deakin University.

Stanley, T.D. and Doucouliagos, C. (H.) (2010) Picture this: A simple graph that reveals much ado about research, *Journal of Economic Surveys*, 24: 170–91.

Stanley, T.D. and Doucouliagos, C. (H.) (2011) Meta-regression approximations to reduce publication selection bias. School Working Paper, Economics Series 2011–4, Deakin University.

Stanley, T.D. and Jarrell, S.B. (1989) Meta-regression analysis: A quantitative method of literature surveys, *Journal of Economic Surveys*, 3: 161–70.

Stanley, T.D. and Jarrell, S.B. (1998) Gender wage discrimination bias? A meta-regression analysis, *Journal of Human Resources*, 33: 947–73.

Stanley, T.D. and Rosenberger, R.S. (2009) Are recreation values systematically underestimated? Reducing publication selection bias for benefit transfer, *Bulletin of Economics and Meta-Analysis*. *http://www.hendrix.edu/maer-network/default.aspx?id=15206*

Stanley, T.D., Doucouliagos, C. (H.) and Jarrell, S.B. (2008) Meta-regression analysis as the socio-economics of economics research, *Journal of Socio-Economics*, 37: 276–92.

Stanley, T.D., Jarrell, S.B. and Doucouliagos, C. (H.) (2010) Could it be better to discard 90% of the data? A statistical paradox, *American Statistician*, 64: 70–7.

Sterling T.D. (1959) Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa, *Journal of the American Statistical Association*, 54: 30–4.

Sterling, T.D., Rosenbaum, W.L. and Weinkam, J.J. (1995) Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa, *American Statistician*, 49: 108–12.

Sterne, J.A. and Egger, M. (2001) Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis, *Journal of Clinical Epidemiology*, 54: 1046–55.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A. and Thacker, S.B. (2000) Meta-analysis of observational studies in epidemiology: A proposal for reporting, *Journal of American Medical Association*, 283: 2008–12.

Sutton, A.J and Higgins, J.P.T. (2007) Recent developments in meta-analysis, *Statistics in Medicine*, 27: 625–50.

Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A. and Song, F. (2000) *Methods for Meta-analysis in Medical Research*, Chichester: Wiley.

Tellis, G.J. (1988) The price elasticity of selective demand: A meta-analysis of econometric models of sales, *Journal of Marketing Research*, 25: 331–41.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms, *Educational Researcher*, 25: 26–30.

Thompson, B. (2004) The "significance crisis" in psychology and education, *Journal of Socio-Economics*, 33: 607–13.

Todorovic, Z.W. and Ma, J. (2008) A review of minimum wage regulation effect: The resource-based view perspective, *Journal of Collective Negotiations*, 32: 57–75.

Tosi, H.L., Werner, S., Katz, J.P. and Gomez-Mejia, L.R. (2000) How much does performance matter? *Journal of Management*, 26: 301–39.

Tullock, G. (1959) Publication decisions and tests of significance – A comment, *Journal of the American Statistical Association*, 54: 593.

United State Environmental Protection Agency (2010) Valuing Mortality Risk Reductions for Environmental Policy: A white paper. Science Advisory Board–Environmental Economics Review Draft.

Valentine, T.J. (1979) Hypothesis tests and confidence intervals for mean elasticities calculated from linear regression equations, *Economics Letters*, 4: 363–67.

Van Houwelingen, H.C., Arends, L.R. and Stijnen, T. (2002) Advanced methods in meta-analysis: Multivariate approach and meta-regression, *Statistics in Medicine*, 21: 589–624.

Verlegh, P.W.J. and Steenkamp, J.E.B.M. (1999) A review and meta-analysis of country-of-origin research, *Journal of Economic Psychology*, 20: 521–46.

Viscusi, W.K. (1993) The value of risks to life and health, *Journal of Economic Literature*, 31: 1912–46.

Viscusi, W.K. and Aldy, J.E. (2003) The value of a statistical life: A critical review of market estimates throughout the world, *Journal of Risk and Uncertainty*, 27: 5–76.

Vista, A.B. and Rosenberger, R. 2009. Primary study aggregation effects: Meta-analysis of sportfishing values in North America. Paper presented at the Oregon State University MAER Network Colloquium, October 1–4 2009, Corvallis.

Wagenaar, A.C., Salois, M.J. and Komro, K.A. (2009) Effects of beverage alcohol price and tax levels on drinking: A meta-analysis of 1003 estimates from 112 studies, *Addiction*, 104: 179–190.

Waldorf, B. and Byun, P. (2005) Meta-analysis of the impact of age structure on fertility, *Journal of Population Economics*, 18: 15–40.

Weichselbaumer, D. and Winter-Ebmer, R. (2005) A meta-analysis of the international gender wage gap, *Journal of Economic Surveys*, 19: 479–511.

Welkowitz, J., Ewen, R.B. and Cohen, J. (1982) *Introductory Statistics for the Behavioral Sciences*, San Diego, CA: Harcourt Brace Jovanovich.

Whitehead, A. (2002) *Meta-Analysis of Controlled Clinical Trials*, Chichester: Wiley.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Wooldridge, J.M. (2006) *Introductory Econometrics: A Modern Approach*, Cincinnati: South-Western.

Wooster, R.B. and Diebel, D.S. (2010) Productivity spillovers from foreign direct investment in developing countries: A meta-regression analysis, *Review of Development Economics*, 14: 640–55.

Young, N.S., Ioannidis, J.P.A. and Al-Ubaydli, O. (2008) Why current publication practices may distort science, *PLoS Med*, 5: doi:10.1371/journal.pmed.0050201.

Zelmer, J. (2003) Linear public goods experiments: A meta-analysis, *Experimental Economics*, 6: 299–310.

Ziliak, S.T. and McCloskey, D.N. (2004) Size matters: The standard error of regressions in the *American Economic Review, The Journal of Socio-Economics*, 33: 527–46.

Index