Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

Dissertation Thesis

Six Essays on Meta-Regression Analysis

Author: **PhDr. Zuzana Havránková**

Advisor: **Prof. Ing. Oldřich Dědek, CSc.**

Academic Year: **2014/2015**

Acknowledgment

I am grateful to my husband, Tomáš Havránek, and to my advisor, Prof. Oldřich Dědek, for help and encouragement, and my children, Kristína and Daniel, for patience. A substantial part of the dissertation was written when I was working on project P402/11/0948 of the Czech Science Foundation, and I thank Prof. Karel Janda for coordinating the project. I am also grateful to the University of California, Berkeley, for hospitality during my research visit in 2014.

Typeset in LaTeX 2_ε using the IES Thesis Template.

Bibliographic Record

Havránková, Zuzana: *Six Essays on Meta-Regression Analysis.* Dissertation thesis. Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies. January 2015, pages 287.

Contents

Abstract

This dissertation thesis consists of six papers on macroeconomics, international economics, and energy economics. All the papers are tied together by the use of meta-regression analysis, which is essential for the derivation of robust policy-relevant conclusions from often conflicting results presented in the empirical literature. I use meta-analysis to quantitatively synthesize the reported research results on a given topic, correct the literature for publication selection bias, and filter out the effect of various misspecifications present in some primary studies. My results can be summarized as follows:

1) The elasticity of intertemporal substitution in consumption, a key input to all dynamic models in finance and macroeconomics, varies significantly across countries. The differences can be explained by the level of stock market participation, when countries with higher participation exhibit larger values of the elasticity; the mean reported elasticity is 0.5. 2) The effect of borders on international trade, which most authors find to be surprisingly large, can be explained away by innovations in methodology introduced in the last decade. When these innovations are taken into account jointly, the border effect disappears for developed countries, and is relatively small for developing countries.

3) When all published estimates of the effect of foreign investment on local firms in the same industries are considered and corrected for publication bias, the literature indicates a zero effect. 4) Publication bias is present also in the literature estimating the effect of foreign investment on local firms in different industries, but here the corrected effect is positive and large. 5) The mean reported price elasticity of gasoline demand is exaggerated twofold due to publication bias. 6) Finally, I also find that publication bias distorts the literature estimating the social cost of carbon emissions, because researchers tend to preferentially report large estimates.

List of Tables

List of Figures

Chapter 1

Introduction

In my dissertation thesis I include six representative examples of my work on meta-analysis in economics. The first paper, "Do Borders Really Slash Trade? A Meta-Analysis," is my latest manuscript and I believe it reflects the state of the art in the field. In contrast, the second paper, "Meta-Analysis of Intra-Industry Spillovers: Update Evidence," is my first meta-analysis, published in 2010. The differences between these two papers demonstrate the development of meta-regression methods in recent years and also the development of my thinking about meta-analysis during my doctoral studies. In the latter part of the Introduction I summarize the results of these two chapters, as well as the results of the other four chapters included in the dissertation. At the end of the Introduction I discuss the evolution of meta-analysis methods and my take on best-practice approaches in the field. Before I proceed to the discussion of the individual papers, I briefly describe my research that I do not include in the dissertation. Many of these papers are also meta-analyses, but I prefer to keep the dissertation relatively short.

I have co-authored thirteen papers during my doctoral studies, seven of which as the corresponding author, and ten of these papers have already been published in peer-reviewed journals. (I use my maiden name, Zuzana Irsova, in all research publications.) My research can be divided into three broad categories. The first category is banking efficiency, where I especially focus on Central and Eastern Europe and apply stochastic frontier analysis and data envelopment analysis to compute efficiency scores of banks and examine their determinants. I have published four papers in this area. The second category is international macroeconomics, especially FDI and trade. I have co-authored seven papers on these topics, and some of them were published in leading field journals: Journal of International Economics, World Development, and Journal of Development Studies. For my work in this area I received the Medal for Research on Development by the Global Development Network (2010), third place in the Young Economist Award by the Czech Economic Society (2010 and 2011), and the Economic Research Award by the Czech National Bank (2012).

The third category of my research interests is energy and environmental economics. Although I have only published one paper in this area ("Demand for Gasoline is More Price-Inelastic than Commonly Thought" in Energy Economics), I currently focus on this field and am involved in three projects in energy and environmental economics (meta-analyses of the social cost of carbon emissions and of climate sensitivity). For my work in environmental and international economics I received the 2013 Award for the Best Students and Graduates in the Czech Republic by the Czech Ministry of Education.

The six papers included in this dissertation span several fields, but they are linked together by the methodology of meta-regression analysis (the papers are co-authored; I assess my contribution in the whole thesis to be roughly 60%). Chapter 2 of this dissertation thesis focuses on the effect of international borders on trade, the paper is co-authored with Tomas Havranek (Havranek & Irsova 2014). The paper is currently at the revise and resubmit stage at the Journal of International Economics. The finding that international borders significantly reduce trade, first reported by McCallum (1995), has become a stylized fact of international economics. A high ratio of trade within national borders to trade across borders, after controlling for other trade determinants, implies large unobserved border barriers, an implausibly high elasticity of substitution between domestic and foreign goods, or both. Obstfeld & Rogoff (2001) include the border effect among the six major puzzles in international macroeconomics, and dozens of researchers have attempted to shrink McCallum's original estimates.

Researchers have proposed several methodological solutions to the border puzzle, such as the inclusion of multilateral resistance terms, consistent measurement of within and between-country distance, and use of disaggregated data. But the border effects reported in the literature are, on average, still close to those estimated by McCallum (1995): regions are likely to trade with foreign regions about fifteen times less than with regions in the same country. The reported border effects do not diminish in time and do not converge to a consensus value that could be used for calibrations. Our goal in this paper is to collect the empirical estimates of the border effect, examine why they vary, and compute a benchmark value for different regions conditional on the implementation of major innovations in the gravity equation. That is, using previously reported results we construct a large synthetic study that estimates the border effect, but corrects for potential publication or misspecification biases.

We collect 32 aspects of studies, such as the characteristics of data, estimation, inclusion of control variables, number of citations, and information on the publication outlet. To explore how these characteristics affect the estimates of the border effect, we employ Bayesian model averaging (Raftery *et al.* 1997). The method addresses model uncertainty inherent in meta-analysis by estimating regressions comprising the potential subsets of the study aspects and weighting them by statistics related to the goodness of fit.

Our results suggest that many innovations in estimating the gravity equation systematically affect the reported border effect: for example, the use of disaggregated data, consistent measure of within and between-country distance, data on actual road or sea distance instead of the great-circle distance, control for multilateral resistance, and the use of the Poisson pseudo-maximum likelihood estimator. When we put these influences together and compute a general equilibrium impact of borders conditional on best practice methodology, we find that borders reduce international trade by only 28% worldwide. The border effects differ significantly across regions—we obtain large estimates for developing and transition countries, but estimates close to zero for most

OECD countries.

We find little evidence of publication bias in the literature: researchers do not preferentially report positive or statistically significant estimates of the border effect. This result is remarkable considering a recent survey of estimates of publication bias, Doucouliagos & Stanley (2013), who show that the problem of selecting intuitive and statistically significant estimates concerns most fields of empirical economics. For example, Ashenfelter *et al.* (1999) find evidence of publication bias in the literature on the returns from schooling, Görg & Strobl (2001) in the estimates of foreign direct investment spillovers, and Rusnak *et al.* (2013) in the literature on the transmission of monetary policy shocks to prices. Unlike many other important parameters in economics, it is easy for researchers to obtain statistically significant estimates of the border effect, so there is little motivation for publication selection.

Chapter 3 presents a meta-analysis of the empirical literature on horizontal productivity spillovers from FDI, it is a joint work with Tomas Havranek (Havránek & Iršová 2010). The paper was published in the Czech Journal of Economics and Finance. We gather a sample of 97 models from 67 studies published either in academic journals or as working papers. Using the vote-counting method, we find that the spillover effect does not seem to be statistically significant in general; employing the approach of Djankov & Murrell (2002), on the other hand, we find evidence that positive spillovers from FDI might exist.

Nevertheless, this is not the case of the narrower sample of studies that were published in the best economics journals or that use panel and firm-level data (and thus are more reliable)—their combined t statistics is insignificant almost in any case. Once publication selection bias is accounted for, the aggregated effect is insignificant, no matter what methodology is used. Therefore, we argue that there is no persuasive empirical evidence on intra-industry spillovers. If there are any horizontal spillover effects, their signs and magnitudes vary from country to country and from industry to industry (we also find some evidence that employment-intensive foreign direct investment may generate relatively higher spillovers through labor turnover).

We further investigate which study aspects affect the reported significance and polarity of spillovers. Nevertheless, we use not only the standard ordinary least squares meta-regression (like Görg & Strobl 2001) but we also employ robust methods (iteratively re-weighted least squares and median regression) as well as pseudo-panel data methods (Meyer & Sinani 2009) and probability models (Wooster & Diebel 2006). Subject to several sensitivity checks we find that, in general, study results are predictably affected by its design, namely by the usage of cross-sectional or panel data, industry- or firm-level aggregation, and specification of the proxy of foreign presence in the industry. Our results suggest that cross-sectional studies tend to report excessively high spillovers, as well as models with industry-level aggregation and employment as a proxy for foreign presence do. However, this pattern appears to become weaker over time, suggesting that newer studies may suffer from such a bias less.

Following Card & Krueger (1995), we test for publication bias in the spillover literature. Contrary to Görg & Strobl (2001), we do not find evidence of publication bias employing this methodology. When the preferred funnel asymmetry test (Doucouliagos & Stanley 2009) is used, however, moderate publication bias is identified in the literature.

Many man-hours of economics and business researchers all over the world have been devoted to investigate horizontal spillovers from foreign direct investment. Is it all "much ado about nothing" as Görg & Greenaway (2004) suggest in the title of their article? While the spillover effect is probably heterogeneous across different countries and industries, the worrying issue is that the results are *systematically* dependent on the chosen methodology. In other words, researchers can influence their results by simply choosing a particular methodology. A strong consensus has formed in the international research community that firm-level panel data are the appropriate tool to test the presence of spillovers from foreign direct investment. For many countries, however, such detailed data are often difficult to construct, and cross-sectional studies are still being published. The outcome of such studies is predictable to a large extent.

The pattern, however, does not concern only the nature of the data. Contrary to Görg & Strobl (2001), our meta-regression analysis shows that the definition of the proxy for foreign presence is important as well and can also bring predictable results. Unfortunately, many studies do not report sensitivity analysis with respect to the definition of foreign presence. When they do, as for instance Geršl (2008), they often find that the spillover effect is not robust. Such pattern of predictability is widespread in economics research. Indeed, Stanley (2001) shows how one of his older meta-regression analyses on the union wage premium (Jarrell & Stanley 1990), coincidentally published in the same issue as a new empirical study on the topic, precisely estimated the results of that study once its characteristics were plugged into the meta-regression. It is natural that heterogeneous research brings heterogeneous results. Researchers should, however, be aware of the predictability pattern, best identified by meta-regression analyses, and report thorough robustness checks.

Chapter 4 focuses on the price elasticity of gasoline demand, it is a joint work with Karel Janda and Tomas Havranek (Havranek *et al.* 2012). The paper was published in Energy Economics. For the purposes of government policy concerning energy security, optimal taxation, and climate change, precise estimates of the price elasticity of gasoline demand are of principal importance. For example, if gasoline demand is highly price-inelastic, taxes will be ineffective in reducing gasoline consumption and the corresponding emissions of greenhouse gases. During the last 30 years the topic has attracted a lot of attention of economists who produced a plethora of empirical estimates of both short- and long-run price elasticities. Yet the estimates vary broadly.

Two international meta-analyses of the elasticity of gasoline demand have been conducted (Espey 1998; Brons *et al.* 2008). These meta-analyses examine carefully the causes of heterogeneity observed in the literature. The average short- and long-run elasticities found by these meta-analyses were -0.26 and -0.58 (Espey 1998) and -0.34 and -0.84 (Brons *et al.* 2008). None of the meta-analyses, however, corrected the estimates for publication selection bias. It is well-known among meta-analysts that publication selection can seriously bias the estimates of price elasticities because positive estimates are usually inconsistent with theory: for instance, Stanley (2005) documents how the price elasticity of water demand is exaggerated *fourfold* because of publication bias.

We employ recently developed meta-analysis methods to test for publication bias and estimate the corrected elasticity beyond. The mixed-effects multilevel meta-regression takes into account heteroscedasticity, which is inevitable in meta-analysis, and between-study heterogeneity, which is likely to occur in most areas of empirical economics. We do not, however, investigate heterogeneity explicitly, as this issue was thoroughly examined by the two previous meta-analyses.

In contrast to previous meta-analyses on this topic, we take into account publication selection bias using the mixed-effects multilevel meta-regression. Publication bias in this area is strong; when we correct for the bias, we obtain estimates of short- and long-run elasticities that are approximately *half*, compared to the results of the previously published meta-analyses and also to the simple mean of all estimates in our sample of literature. If the simple mean reflects our profession's impression about the magnitude of the price elasticity of gasoline demand, the impression exaggerates the true elasticity twofold.

This paper complements the previously published meta-analyses on the price elasticity of gasoline demand (Espey 1998; Brons *et al.* 2008). These meta-analyses focus on the reasons why estimates of elasticities differ for different regions and different methods used and provide mean estimates of short- and long-run price elasticities as a bonus. It is important to bear in mind the differences between the methods used in this paper to deliver the average estimates of elasticity and the methods used in Espey (1998) and Brons *et al.* (2008). First, the estimates of Brons *et al.* (2008) are based on a seemingly unrelated regression model with cross-equation restrictions. Second, neither Espey (1998) nor Brons *et al.* (2008) use a multilevel approach to distinguish between study-level and estimate-level variation. Third, the sets of studies differ among the three meta-analyses. Although the estimates of average elasticity are therefore not directly comparable, we argue there is a strong case for the presence of publication bias in favor of larger negative estimates of elasticities in the literature.

The estimated elasticities corrected for publication bias, -0.09 for the short run and -0.31 for the long run, are average across many countries, methods, and time periods; we report them as reference values. A similar pattern of publication bias, however, is likely to appear in any subset of the literature. Thus large negative estimates of price elasticities should be taken with a grain of salt.

Chapter 5 focuses on spillovers from foreign direct investment to local firms, it is a joint work with Tomas Havranek (Havranek & Irsova 2012). The paper was published in the Journal of Development Studies. Policy makers, especially in transition and developing countries, usually encourage inward FDI in expectation that domestic firms in the same sectors benefit from know-how brought by foreigner investors. Moreover, many of such policy makers believe that firms in supplier sectors benefit from direct knowledge transfers from foreigners, and perhaps also that firms in customer sectors benefit from higher-quality intermediate inputs produced by foreigners. With an allusion to the production chain, the effect of foreign presence on the productivity of domestic competitors is typically labeled *horizontal spillovers*, the effect on domestic suppliers *backward spillovers*, and the effect on domestic customers *forward spillovers*; backward and forward spillovers together are called vertical spillovers. Although not a necessary nor sufficient condition for the provision of government subsidies for FDI, spillover effects are highly policy-relevant. In consequence, the search for spillovers has given rise to a burgeoning stream of empirical literature in development economics, and we investigate 57 such papers in this meta-analysis.

Horizontal spillovers are usually thought to occur through three main channels. The first channel is the competition effect (for example, Aitken & Harrison 1999): the entry of foreign firms increases competition in the domestic market. Increased competition forces domestic firms to use their inputs more efficiently, boosting their productivity. Nonetheless, increased competition also reduces the opportunities of domestic firms to exploit returns to scale, reducing their productivity. The second channel is the demonstration effect (for example, Blomstrom & Kokko 1998): foreign investors bring technology more advanced than that of domestic firms, especially in transition and developing countries. In this way, foreigners "demonstrate" up-to-date technology to domestic firms, which imitate and implement it. The third channel is labor turnover (for example, Görg & Greenaway 2004): foreign firms train local employees, who accumulate know-how and experience with modern technology. Eventually, locals change the employer or start a firm of their own, diffusing knowledge further.

Foreign affiliates will try to prevent the transfer of technology to their competitors; that is, they will try to minimize the positive effects of demonstration and labor turnover. Therefore if the detrimental effects of competition prevail, horizontal spillovers altogether may well become insignificant or even negative. On the other hand, foreigners have incentives to provide assistance to their local suppliers, since they want to ensure a high quality and on-time delivery of inputs.

Therefore, the recent literature (Javorcik 2004; Blalock & Gertler 2008) emphasizes vertical linkages between foreign investors and domestic firms. The per-job value of spillovers stirred up by linkages can be compared with the amount of government subsidies, as Haskel *et al.* (2007) do; hence for policy recommendations precise estimates of spillovers are required. Since the results of individual studies vary broadly, a quantitative literature survey, meta-analysis, represents a useful method to obtain robust estimates of spillovers (Stanley 2001). If, however, some particular results are more likely to be published (for example, those consistent with the mainstream intuition about spillovers outlined in the paragraphs above), a simple average of the reported results will be a biased estimate of the underlying spillover effect. The importance of publication selection bias in the spillover literature was stressed already by the first meta-analysis on this topic, Görg & Strobl (2001).

In contrast to the earlier meta-analyses on FDI spillovers (Görg & Strobl 2001; Meyer & Sinani 2009), we examine backward and forward spillovers in addition to horizontal spillovers. Using a large data set, we employ modern meta-analysis methods developed by Stanley (2005; 2008) to estimate the underlying spillover effects and the magnitude of publication bias. We present individual surveys for each country inspected in the literature and construct a unique cross-country data set of estimated spillovers. Furthermore, we retrieve estimates of publication bias for each study and examine how the intensity of publication selection depends on the characteristics of the authors, such as affiliation, experience, and tenure pressure.

Our results suggest that the average effect of foreign affiliates on the productivity of their local competitors (horizontal spillover) is economically insignificant. The effect of foreign affiliates on their local customers (forward spillover) is likewise negligible. On the other hand, we detect a statistically significant and economically meaningful effect of foreign affiliates on their local suppliers (backward spillover). Specifically, a 10-percentage-point increase in foreign presence is associated with a 1.2% boost to the productivity of domestic firms in supplier sectors. Such a spillover effect is consistent with subsidies for FDI. Nevertheless, policy makers should exercise caution because the estimates capture more than externalities: studies on FDI spillovers do not account for possible compensations for the transfer of technology (Keller 2009). An exception is Blalock & Gertler (2008), who examine the influence of foreign presence on the profits of Indonesian firms and confirm the positive externality.

While the average backward spillover is robustly positive, it differs significantly across countries. For example, the effect for all developing countries examined by the studies in our sample is twice as large as the average spillover reported for developed countries. The degree of economic development plays an important role in explaining the difference, but it is not the only one. In a companion paper (Havranek & Irsova 2011) we examine in detail what causes the differences in the reported FDI spillovers. We find that both the characteristics of the host country and the characteristics of FDI matter. For example, a larger technology gap of domestic firms with respect to foreign investors is associated with less spillovers. On the other hand, a higher degree of trade openness is associated with more spillovers from inward FDI. The mode of entry of FDI is also important: fully foreign-owned investments generate less positive spillovers than joint projects of foreign and domestic firms. In the present paper we take stock of the empirical research on FDI spillovers and provide a unique database of average estimates for each country examined in the literature.

Chapter 6 focuses on the cross-country heterogeneity in intertemporal substitution, it is a joint work with Roman Horvath, Tomas Havranek, and Marek Rusnak

(Havranek *et al.* 2013). The paper is currently at the revise and resubmit stage at the Journal of International Economics. The elasticity of intertemporal substitution in consumption (EIS) reflects households' willingness to substitute consumption between time periods in response to changes in the expected real interest rate. Therefore it represents a crucial parameter for a wide range of economic models involving intertemporal choice, from modeling the behavior of aggregate savings and the impact of fiscal policy to computing the social cost of carbon emissions, and has been estimated by hundreds of researchers.

The most cited empirical study estimating the elasticity, Hall (1988), who concludes that the EIS is not likely to be larger than 0.1, has influenced many researchers. Some studies use a value of 0.2 (Chari *et al.* 2002; House & Shapiro 2006; Piazzesi *et al.* 2007), or a value of 0.5 (Jin 2012; Trabandt & Uhlig 2011; Rudebusch & Swanson 2012), or a value of 2 (Ai 2010; Barro 2009; Colacito & Croce 2011), to name but a few recent examples of different calibrations. The reason for the different calibrations is differences in the results of empirical studies on the EIS. For example, the standard deviation of the estimates reported by the 33 studies in our sample which were published in the top five general interest journals is 1.4, outliers excluded. Most commentators would agree with Ai (2010, p. 1357), who starts his discussion of calibration by noting that "empirical evidence on the magnitude of the EIS parameter is mixed."

In this paper we collect 2,735 estimates of the elasticity of intertemporal substitution reported in 169 studies and review the literature quantitatively using meta-analysis methods. While controlling for differences in methodology, we focus on explaining country-level heterogeneity. The studies in our sample provide us with estimates of the EIS for 104 countries, and we show that the mean values reported for the countries vary substantially. We build on the literature that explores the heterogeneity in the EIS at the micro level. For example, Blundell *et al.* (1994) and Attanasio & Browning (1995) suggest that rich households tend to show a larger elasticity of intertemporal substitution, and we examine whether GDP per capita is associated with the mean EIS reported for the country. Mankiw & Zeldes (1991) and Vissing-Jorgensen (2002) find a larger elasticity for stockholders than for non-stockholders, and we explore the relationship between stock market participation and the elasticity of intertemporal substitution at the country level. Bayoumi (1993) and Wirjanto (1995), among others, indicate that liquidity-constrained households show a smaller EIS, and we examine whether ease of access to credit helps explain the cross-country variation in the elasticity.

The mean estimate of the elasticity of intertemporal substitution reported in empirical studies is 0.5, but we show that cross-country differences are important. Since it is often unclear which aspects of methodology should matter for the magnitude of the estimated EIS, we include all 30 that we collect and employ Bayesian model averaging (Raftery *et al.* 1997) to deal with the resulting model uncertainty. Our findings suggest that a larger EIS is associated with higher per capita income of the country, and especially with higher stock market participation. According to our baseline model, a 10-percentage-point increase in the rate of stock market participation is associated with an increase in the EIS of 0.24. Moreover, wealth and asset market participation are also important at the micro level: studies estimating the EIS using a sub-sample of rich households or asset holders find on average an EIS larger by 0.21.

Chapter 7 focuses on social costs of carbon emissions, it is a joint work with Tomas Havranek, Karel Janda, and David Zilberman (Havranek *et al.* 2014). The paper is currently submitted to the Energy Journal. The social cost of carbon (SCC) is a key parameter for the formulation of climate policy. If the SCC was pinned down precisely, policy makers could use the parameter to set the optimal carbon tax. For this reason, dozens of researchers using different families of models have estimated the SCC—but their findings and the resulting policy implications vary greatly. Several previous studies have offered quantitative surveys of the literature (Tol 2005; 2013), focusing especially on the characteristics of study design that may influence the reported estimates, but no study has discussed nor tested for the potential selective reporting bias in the estimates of the social cost of carbon.

Several studies examine selective reporting in the context of climate change research. The problem is widely discussed in phenology (Both *et al.* 2004; Gienapp *et al.* 2007; Menzel *et al.* 2006), and the evidence suggests that while selective reporting is a minor issue in multi-species studies, positive results from single-species studies are reported more often than neutral results (Parmesan 2007). Maclean & Wilson (2011) conduct a meta-analysis of the relation between climate change and extinction risk and find mixed results concerning selective reporting, with evidence for the bias among estimates of extinction risk, but no bias among estimates of high extinction risk. Michaels (2008) examines 166 papers on climate change published in Science and Nature and argues that there is substantial evidence for selective reporting. Swanson (2013) indicates that many of the current model simulations of climate change are inconsistent with the observed changes in air temperature and the frequency of monthly temperature extremes, which might be due to selective reporting. In contrast, Darling & Côté (2008) investigate the relationship between climate change and biodiversity loss and find no evidence of selective reporting, and Massad & Dyer (2010) find no signs of selective reporting in the literature on the effects of climate change on plant-herbivore interactions.

In contrast to most subjects of meta-analysis in economics, the SCC is not estimated in a regression network. Rather, it is a result of a complex calibration exercise, and the uncertainty surrounding the estimates is usually determined via Monte Carlo simulations. Therefore the literature lacks the usual suspects when it comes to potential selective reporting: specification search across models with different control variables, choice of the estimation technique, and the selection of the data sample. On the other hand, the authors have the liberty to choose among many possible values of the parameters that enter the computation and influence both the estimated magnitude of the SCC and the associated uncertainty. In a critical review of integrated assessment models, Pindyck (2013, p. 863) argues that "these models can be used to obtain almost any result one desires." Despite the difficulty in computing the SCC, we believe it is worth trying to pin down this crucial parameter. Testing for the potential selective reporting bias represents a part of this effort.

We examine 809 estimates of the SCC reported in 101 primary studies. We employ meta-regression methods commonly used in economics and other fields to detect potential selective reporting in the literature. Our results suggest that, on average, the authors of primary studies tend to report preferentially estimates for which the 95% confidence interval excludes zero, which creates an upward bias in the literature. In other words, we observe that small estimates of the SCC are associated with less uncertainty (expressed as the approximate standard error used to compute the lower bound of the confidence interval) than large estimates. The finding suggests that some small estimates with large uncertainty—that is, not ruling out negative values of the SCC—might be selectively omitted from the literature. Our results also indicate that selective reporting tends to be stronger in studies published in peer-reviewed journals than in unpublished manuscripts.

Lessons Learned

The papers included in the dissertation and summarized on the previous pages use many different meta-analysis methods, which might puzzle the reader. One of the reasons for the differences is the time when these papers where published, which reflects the evolution of meta-regression methods; sometimes, however, the choice of a particular meta-analysis technique depends on the specific data set or research question under examination. Up-to-date guidelines for conducting meta-regression analysis in economics are provided by Stanley *et al.* (2013), but I still consider it useful to briefly summarize my take on best-practice methods in the field and provide practical details that are missing in other guidelines. I structure the discussion into several paragraphs according to the issues that meta-analysts face.

Selection of primary studies The first problem that a meta-analyst faces is which studies to include in the meta-analysis. The typical recommendation (Stanley 2001) is to use all studies estimating the parameter in question, if possible. Sometimes, however, such an approach is not feasible because hundreds or even thousands of papers exist on the topic. In this dissertation it is the case of, for example, the literature estimating the elasticity of intertemporal substitution in consumption. Rather than selecting a random sample of studies, I argue it makes sense to focus on published studies only. Published studies can be, ex ante, expected to be of higher quality, are better typeset (which makes data collection easier and reduces the danger of typos), and unlike working papers, for published studies there is only one version available (which makes it easier to date the study). Moreover, several meta-analyses show that there is little difference between published and unpublished studies in the extent of publication bias (Rusnak *et al.* 2013).

Tests of publication bias I prefer the funnel asymmetry test discussed by Stanley (2005), because it has been shown to perform well in Monte Carlo simulations and is very intuitive. The test is based on the realization that in the absence of publication bias there should be no systematic relation between estimates and their standard errors. The authors of primary studies usually report t-statistics for their estimates, which means that they assume that the ratio of the estimates to their standard errors have a t-distribution, which in turn implies that estimates and standard errors should be statistically independent quantities. If, on the other hand, researchers prefer to publish estimates with a particular sign or statistical significance, estimates will be correlated with standard errors. The regression is heteroskedastic, so weighted least squares (with inverse of the variance as the weight) should be used. If possible, researchers should use study-level fixed effects and cluster standard errors at the study level. I also recommend to use the inverse of the square root of the number of observations as an instrument for the standard error. If the meta-analyst fails to control for a method choice that affects the estimates and their standard errors in the same direction, he or she obtains biased estimates of the extent of publication bias. The inverse of the square root of the number of observations is usually a valid instrument, because it is obviously correlated with the standard error, but not correlated with most method choices.

Selection of variables Most applications of meta-analysis involve dozens of variables that may potentially affect the magnitude of the parameter in question. It is not clear which variables should be selected in the baseline model, because for many of them we have little theoretical guidance (for example, the effect of the number of observations), but we still want to control for these aspects of study design. I recommend not to use sequential t-tests and remove the least significant variables one by one; such an approach is not statistically valid. Instead, meta-analysts should use Bayesian model averaging, which is a method that formally addresses model uncertainty in meta-analysis. The methods runs millions of regressions with different combinations of all explanatory variables and makes a weighted average over them (with weights being approximately proportional to the goodness of fit of the individual models).

Robustness checks It is a matter of taste whether to use weighted least squares in meta-analysis when other explanatory variables than the standard error are included. Tom Stanley argues to always use weighted least squares, because of the heteroskedasticity problem and because weighting always gives priority to more precise results. I prefer not to weight the regression by precision if the regression contains variables defined on the study level, like the number of citations. Because precision differs for each estimate within a study, weighting by precision introduces artificial variation in these variables. Since both approaches often yield very different results, it might be a good idea to report the results of the other approach as a robustness check. Moreover, if the meta-analyst cannot use study-level fixed effects when estimating publication bias (for example, because many studies report only one estimate), it is advisable to report both simple OLS estimates and mixed-effects estimates (which give each study approximately the same weight even though different studies report a different number of estimates).

Judgment in meta-analysis Although meta-analysis is a formal method of literature surveys, it does not mean that it is judgment-free. I argue that a good meta-analysis should discuss which method choices in the primary studies are preferable and, if possible, it should try to construct an estimate of the mean effect corrected for both publication bias and misspecifications in primary studies. In practical terms, the estimate is derived as a linear combination from the final specification, when the meta-analyst plugs in the preferred values for each variable (for example, "1" for the dummy variable that reflects whether the primary study controls for endogeneity by instrumenting the explanatory variable). Such a "best practice" estimation is often controversial, but I believe it is the principal value added of any meta-analysis.

References

Ai, H. (2010): "Information Quality and Long-Run Risk: Asset Pricing Implications." *Journal of Finance* **65(4)**: pp. 1333–1367.

Aitken, B. J. & A. E. Harrison (1999): "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela." *American Economic Review* **89(3)**: pp. 605–618.

Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* **6(4)**: pp. 453–470.

Attanasio, O. P. & M. Browning (1995): "Consumption over the Life Cycle and over the Business Cycle." *American Economic Review* **85(5)**: pp. 1118–37.

Barro, R. J. (2009): "Rare Disasters, Asset Prices, and Welfare Costs." *American Economic Review* **99(1)**: pp. 243–64.

Bayoumi, T. A. (1993): "Financial Deregulation and Consumption in the United Kingdom." *The Review of Economics and Statistics* **75(3)**: pp. 536–39.

Blalock, G. & P. J. Gertler (2008): "Welfare gains from Foreign Direct Investment through technology transfer to local suppliers." *Journal of International Economics* **74(2)**: pp. 402–421.

Blomstrom, M. & A. Kokko (1998): "Multinational Corporations and Spillovers." *Journal of Economic Surveys* **12(3)**: pp. 247–77.

Blundell, R., M. Browning, & C. Meghir (1994): "Consumer Demand and the Life-Cycle Allocation of Household Expenditures." *Review of Economic Studies* **61(1)**: pp. 57–80.

Both, C., A. V. Artemyev, B. Blaauw, R. J. Cowie, A. J. Dekhuijzen, T. Eeva, A. Enemar, L. Gustafsson, E. V. Ivankina, A. Jarvinen, N. B. Metcalfe, N. E. I. Nyholm, J. Potti, P.-A. Ravussin, J. J. Sanz, B. Silverin, F. M. Slater, L. V. Sokolov, J. Torok, , W. Winkel, J. Wright, H. Zang, & M. E. Visser (2004): "Large-scale geographical variation confirms that climate change causes birds to lay earlier." *Proceedings of the Royal Society of London. Series B, Biological Sciences* **271**: p. 1657–1662.

Brons, M., P. Nijkamp, E. Pels, & P. Rietveld (2008): "A meta-analysis of the price elasticity of gasoline demand. A SUR approach." *Energy Economics* **30(5)**: pp. 2105–2122.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review* **85(2)**: pp. 238–43.

Chari, V. V., P. J. Kehoe, & E. R. McGrattan (2002): "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?" *Review of Economic Studies* **69(3)**: pp. 533–63.

Colacito, R. & M. M. Croce (2011): "Risks for the Long Run and the Real Exchange Rate." *Journal of Political Economy* **119(1)**: pp. 153–181.

Darling, E. S. & I. M. Côté (2008): "Quantifying the evidence for ecological synergies." *Ecology Letters* **11(12)**: pp. 1278–1286.

Djankov, S. & P. Murrell (2002): "Enterprise Restructuring in Transition: A Quantitative Survey." *Journal of Economic Literature* **40(3)**: pp. 739–792.

Doucouliagos, H. & T. D. Stanley (2009): "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* **47(2)**: pp. 406–428.

Doucouliagos, H. & T. D. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Espey, M. (1998): "Gasoline demand revisited: an international meta-analysis of elasticities." *Energy Economics* **20(3)**: pp. 273–295.

Geršl, A. (2008): "Productivity, Export Performance, and Financing of the Czech Corporate Sector: The Effects of Foreign Direct Investment." *Czech Journal of Economics and Finance* **58**: pp. 232–247.

Gienapp, P., R. Leimu, & J. Merila (2007): "Responses to climate change in avian migration time–microevolution versus phenotypic plasticity." *Climate Research* **35**: p. 25–35.

Görg, H. & D. Greenaway (2004): "Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment?" *World Bank Research Observer* **19(2)**: pp. 171–197.

Görg, H. & E. Strobl (2001): "Multinational Companies and Productivity Spillovers: A Meta-analysis." *The Economic Journal* **111(475)**: pp. F723–39.

Hall, R. E. (1988): "Intertemporal Substitution in Consumption." *Journal of Political Economy* **96(2)**: pp. 339–57.

Haskel, J. E., S. C. Pereira, & M. J. Slaughter (2007): "Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?" *The Review of Economics and Statistics* **89(3)**: pp. 482–496.

Havranek, T., R. Horvath, Z. Irsova, & M. Rusnak (2013): "Cross-Country Heterogeneity in Intertemporal Substitution." *William Davidson Institute Working Papers Series wp1056*, William Davidson Institute at the University of Michigan.

Havranek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85(2)**: pp. 234–244.

Havranek, T. & Z. Irsova (2012): "Survey Article: Publication Bias in the Literature on Foreign Direct Investment Spillovers." *Journal of Development Studies* **48(10)**: pp. 1375–1396.

Havranek, T. & Z. Irsova (2014): "Do Borders Really Slash Trade? A Meta-Analysis." *working paper*, Charles University, Prague.

Havranek, T., Z. Irsova, & K. Janda (2012): "Demand for Gasoline is More Price-Inelastic than Commonly Thought." *Energy Economics* **34(1)**: p. 201–207.

Havranek, T., Z. Irsova, K. Janda, & D. Zilberman (2014): "Selective Reporting and the Social Cost of Carbon." *Department of Agricultural & Resource Economics, UC Berkeley, Working Paper Series, Department of Agricultural & Resource Economics, UC Berkeley 1139*, Department of Agricultural & Resource Economics, UC Berkeley.

Havránek, T. & Z. Iršová (2010): "Meta-Analysis of Intra-Industry FDI Spillovers: Updated Evidence." *Czech Journal of Economics and Finance (Finance a uver)* **60(2)**: pp. 151–174.

House, C. L. & M. D. Shapiro (2006): "Phased-In Tax Cuts and Economic Activity." *American Economic Review* **96(5)**: pp. 1835–1849.

Jarrell, S. B. & T. D. Stanley (1990): "A meta-analysis of the union-nonunion wage gap." *Industrial and Labor Relations Review* **44(1)**: pp. 54–67.

Javorcik, B. S. (2004): "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages." *American Economic Review* **94(3)**: pp. 605–627.

Jin, K. (2012): "Industrial Structure and Capital Flows." *American Economic Review* **102(5)**: pp. 2111–46.

Keller, W. (2009): "International Trade, Foreign Direct Investment, and Technology Spillovers." *NBER Working Papers 15442*, National Bureau of Economic Research.

MACLEAN, I. M. D. & R. J. WILSON (2011): "Recent ecological responses to climate change support predictions of high extinction risk." *Proceedings of the National Academy of Sciences of the United States of America* **108(30)**: p. 12337–12342.

MANKIW, N. G. & S. P. ZELDES (1991): "The Consumption of Stockholders and Non-Stockholders." *NBER Working Papers 3402*, National Bureau of Economic Research, Inc.

MASSAD, T. J. & L. A. DYER (2010): "A meta-analysis of the effects of global environmental change on plant-herbivore interactions." *Arthropod-Plant Interactions* **4(3)**: p. 181–188.

MCCALLUM, J. (1995): "National Borders Matter: Canada-U.S. Regional Trade Patterns." *American Economic Review* **85(3)**: pp. 615–23.

MENZEL, A., T. H. SPARKS, N. ESTRELLA, E. KOCH, A. AASA, R. AHAS, K. ALM-KUBLER, P. BISSOLLI, O. BRASLAVSKÁ, A. BRIEDE, F. M. CHMIELEWSKI, Z. CREPINSEK, Y. CURNEL, A. DAHL, C. DEFILA, A. DONNELLY, Y. FILELLA, K. JATCZAK, F. MAGE, A. MESTRE, O. NORDLI, J. PENUELAS, P. PIRINEN, V. REMIŠOVÁ, H. SCHEIFINGER, M. STRIZ, A. SUSNIK, A. J. H. VAN VLIET, F.-E. WIELGOLASKI, S. ZACH, & A. ZUST (2006): "European phenological response to climate change matches the warming pattern." *Global Change Biology* **12(10)**: p. 1969–1976.

MEYER, K. E. & E. SINANI (2009): "When and where does foreign direct investment generate positive spillovers? A meta-analysis." *Journal of International Business Studies* **40(7)**: pp. 1075–1094.

MICHAELS, J. P. (2008): "Evidence for "Publication Bias" Concerning Global Warming in Science and Nature." *Energy & Environment* **19(2)**: pp. 287–301.

OBSTFELD, M. & K. ROGOFF (2001): "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" In "NBER Macroeconomics Annual 2000, Volume 15," NBER Chapters, pp. 339–412. National Bureau of Economic Research, Inc.

Parmesan, C. (2007): "Influences of species, latitudes and methodologies on estimates of phenological response to global warming." *Global Change Biology* **13**: p. 1860–1872.

Piazzesi, M., M. Schneider, & S. Tuzel (2007): "Housing, Consumption and Asset Pricing." *Journal of Financial Economics* **83(3)**: pp. 531–569.

Pindyck, R. S. (2013): "Climate Change Policy: What Do the Models Tell Us?" *Journal of Economic Literature* **51(3)**: pp. 860–72.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* **92**: pp. 179–191.

Rudebusch, G. D. & E. T. Swanson (2012): "The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks." *American Economic Journal: Macroeconomics* **4(1)**: pp. 105–43.

Rusnak, M., T. Havranek, & R. Horvath (2013): "How to Solve the Price Puzzle? A Meta-Analysis." *Journal of Money, Credit and Banking* **45(1)**: pp. 37–70.

Stanley, T., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh, R. S. Rosenberger, & K. Rost (2013): "Meta-Analysis of Economics Research Reporting Guidelines." *Journal of Economic Surveys* **27(2)**: pp. 390–394.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

SWANSON, K. L. (2013): "Emerging selection bias in large-scale climate change simulations." *Geophysical Research Letters* **40(12)**: p. 3184–3188.

TOL, R. S. J. (2005): "The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties." *Energy Policy* **33(16)**: p. 2064–2074.

TOL, R. S. J. (2013): "Targets for global climate policy: An overview." *Journal of Economic Dynamics and Control* **37(5)**: pp. 911–928.

TRABANDT, M. & H. UHLIG (2011): "The Laffer Curve Revisited." *Journal of Monetary Economics* **58(4)**: pp. 305–327.

VISSING-JORGENSEN, A. (2002): "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *Journal of Political Economy* **110(4)**: pp. 825–853.

WIRJANTO, T. S. (1995): "Aggregate Consumption Behaviour and Liquidity Constraints: The Canadian Evidence." *Canadian Journal of Economics* **28(4b)**: pp. 1135–52.

WOOSTER, R. B. & D. S. DIEBEL (2006): "Productivity Spillovers from Foreign Direct Investment in Developing Countries: A Meta-Regression Analysis." *Working paper*, Available at SSRN: http://ssrn.com/abstract=898400.

Chapter 2

Do Borders Really Slash Trade? A Meta-Analysis

Abstract

National borders reduce trade, but most estimates of the border effect seem puzzlingly large. We show that major methodological innovations of the last decade combine to shrink the border effect to a mere 28% reduction in international trade flows worldwide. The border effect varies across regions: it is large in emerging countries, but close to zero in OECD countries. For the computation we collect 1,271 estimates of the border effect reported in 61 studies, codify 32 aspects of study design that may influence the estimates, and use Bayesian model averaging to take into account model uncertainty. Our results suggest that methods systematically affect the estimated border effects. Especially important is the level of aggregation, measurement of internal and external distance, control for multilateral resistance, and treatment of zero trade flows. We find no evidence of publication bias.

Keywords: Bayesian mode averaging, bilateral trade, borders, gravity, meta-analysis, publication selection bias

JEL Codes: F14, F15

This paper is a joint work with Tomas Havranek. We are grateful to Marek Rusnak, Jiri Schwarz, and Diana Zigraiova for their helpful comments. An online appendix with data and code is available at `meta-analysis.cz/border`. The paper is currently under review at the Economic Journal.

2.1 Introduction

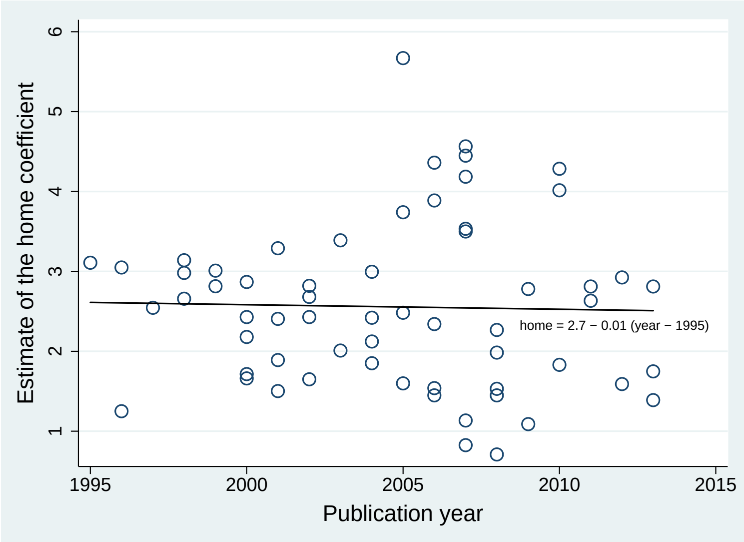
The finding that international borders significantly reduce trade, first reported by McCallum (1995), has become a stylized fact of international economics. A high ratio of trade within national borders to trade across borders, after controlling for other trade determinants, implies large unobserved border barriers, an implausibly high elasticity of substitution between domestic and foreign goods, or both. Obstfeld & Rogoff (2001) include the border effect among the six major puzzles in international macroeconomics, and dozens of researchers have attempted to shrink McCallum's original estimates.

Researchers have proposed several methodological solutions to the border puzzle, such as the inclusion of multilateral resistance terms, consistent measurement of within and between-country distance, and use of disaggregated data. But the border effects reported in the literature are, on average, still close to those estimated by McCallum (1995): regions are likely to trade with foreign regions about fifteen times less than with regions in the same country.

Figure 2.1 shows that new methods and data sets used in the gravity equation, the workhorse tool for computing border effects, increase the dispersion of results. The reported border effects do not diminish in time and do not converge to a consensus value that could be used for calibrations. Our goal in this paper is to collect the empirical estimates of the border effect, examine why they vary, and compute a benchmark value for different regions conditional on the implementation of major innovations in the gravity equation. That is, using previously reported results we construct a large synthetic study that estimates the border effect, but corrects for potential publication or misspecification biases.

We employ the framework of meta-analysis, the quantitative method of research synthesis (Stanley 2001). Meta-analysis has been used in economics by, for instance, Card & Krueger (1995) on the employment effects of minimum wage increases, Disdier & Head (2008) on the impact of distance on trade, Havranek & Irsova (2011) on the relation between foreign investment and local firms' productivity, and Chetty

Figure 2.1: The reported border effects diverge, not decrease



Notes: The figure depicts median estimates of the "home coefficient" (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows) reported in individual studies. The border effect can be obtained by exponentiating the home coefficient: the mean is $\exp(2.7) = 15$. The horizontal axis measures the year when the first drafts of studies appeared in Google Scholar. The black line shows linear fit.

et al. (2011) on the intertemporal elasticity of substitution in labor supply. We collect 32 aspects of studies, such as the characteristics of data, estimation, inclusion of control variables, number of citations, and information on the publication outlet. To explore how these characteristics affect the estimates of the border effect, we employ Bayesian model averaging (Raftery *et al.* 1997). The method addresses model uncertainty inherent in meta-analysis by estimating regressions comprising the potential subsets of the study aspects and weighting them by statistics related to the goodness of fit.

Our results suggest that many innovations in estimating the gravity equation systematically affect the reported border effect: for example, the use of disaggregated data, consistent measure of within and between-country distance, data on actual road or sea distance instead of the great-circle distance, control for multilateral resistance, and the use of the Poisson pseudo-maximum likelihood estimator. When we put these influences together and compute a general equilibrium impact of borders conditional on best practice methodology, we find that borders reduce international trade by only 28% worldwide. The border effects differ significantly across regions—we obtain large estimates for developing and transition countries, but estimates close to zero for most OECD countries.

We find little evidence of publication bias in the literature: researchers do not preferentially report positive or statistically significant estimates of the border effect. This result is remarkable considering a recent survey of estimates of publication bias, Doucouliagos & Stanley (2013), who show that the problem of selecting intuitive and statistically significant estimates concerns most fields of empirical economics. For example, Ashenfelter *et al.* (1999) find evidence of publication bias in the literature on the returns from schooling, Görg & Strobl (2001) in the estimates of foreign direct investment spillovers, and Rusnak *et al.* (2013) in the literature on the transmission of monetary policy shocks to prices. Unlike many other important parameters in economics, it is easy for researchers to obtain statistically significant estimates of the border effect, so there is little motivation for publication selection.

The remainder of the paper is organized as follows. Section 2.2 describes how we collect data from studies and discusses the basic properties of the data set. Section 2.3 tests for publication bias in the literature. Section 2.4 explores the heterogeneity in the estimated border effects and constructs best practice estimates for different regions. Section 2.5 presents robustness checks. Section 2.6 discusses the potential pitfalls of meta-analysis. Section 2.7 concludes. Appendix A presents diagnostics of Bayesian model averaging, and the online appendix at `meta-analysis.cz/border` provides the data and code we use in the paper.

2.2 The Border Effects Data Set

The studies from which we collect estimates of the border effect assume that trade flows are generated by the following general definition of the gravity equation:

$$\text{Trade}_{ij} = G \cdot \text{Exporter}_i \cdot \text{Importer}_j \cdot \text{Distance}_{ij}^{-\alpha} \cdot \exp(home \cdot \text{Same country}_{ij}) \cdot \text{Access}_{ij}, \tag{2.1}$$

where $Trade_{ij}$ denotes the volume of trade flows from region i to region j, G is a "gravitational" constant, $Exporter_i$ denotes exporting capabilities of region i with respect to all trading partners, $Importer_j$ denotes the characteristics of region j that affect imports from all trading partners, $Distance_{ij}$ denotes the distance between regions i and j, $Same\ country_{ij}$ denotes a dummy variable that equals one if regions i and j belong to the same country, and $Access_{ij}$ denotes all other bilateral accessibility characteristics between regions i and j.

The authors usually estimate a log-linearized version of (2.1) with exporter and importer fixed effects to control for multilateral resistance terms. Some authors use non-linear estimators, and even for the linear estimation there are many method choices the authors must make. We identify 32 aspects of study design that may potentially influence the estimate of the border effect and explain them in detail in Section 2.4. We collect estimates of *home* reported in studies, which is the semi-elasticity corresponding to the ratio of within to between-country trade flows; the border effect can be obtained by exponentiating the semi-elasticity. It is convenient to analyze the semi-elasticities because authors provide standard errors for them and the estimates should be approximately normally distributed.

Our data sources are studies that estimate the semi-elasticities; we call them primary studies and search for them using the RePEc database. We use the following search query for titles, keywords, and abstracts of papers listed in the database: `(border OR home bias) AND trade AND gravity`. The search yields 370 hits since 1995. We read the abstracts of all the studies and download those that show a promise of containing empirical estimates of the border effect. Additionally we examine the references of the studies and obtain other papers that might provide empirical estimates. We stop the search on January 1, 2014. The list of all examined studies is available in the online appendix at `meta-analysis.cz/border`.

We apply three inclusion criteria. First, the study must examine the effect of international borders. That is, we exclude studies estimating intranational border effects (for example, Wolf 2000). We expect the mechanism driving borders effects in intranational trade to be different enough to call for a separate meta-analysis.

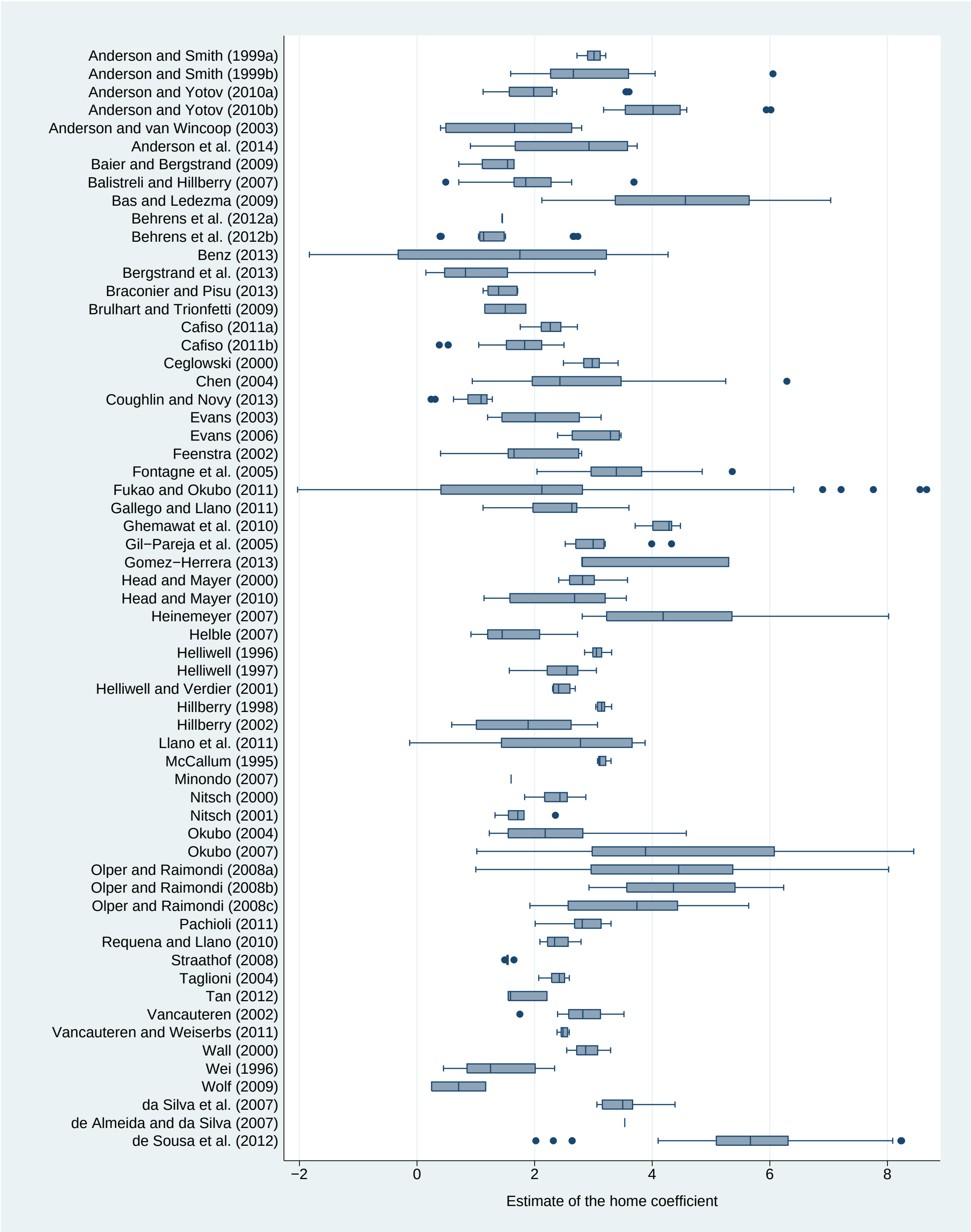
Second, we exclude papers that include the "same nation" dummy in the gravity equation as a control variable for territories, such as the overseas departments of France (for example, Rose 2000). The "same nation" dummy has little variation and often captures trade between a large country and its small territories. Third, we only include studies that provide standard errors for their estimates—or statistics from which standard errors can be computed. Without estimates of standard errors we cannot test for publication bias in the literature. While we conduct the search using English keywords, we do not further exclude any studies based on the language of publication.

The 61 studies that conform to our selection criteria are listed in the online appendix. Of these, 48 are published in refereed journals and 13 are working papers or mimeographs; later in the analysis we control for the publication outlet of the study and other aspects of quality. The median study in our sample was published in 2007, which shows that the literature estimating the border effects is alive and well, with more and more studies coming out each year. Together the studies have received almost 11,000 citations in Google Scholar, or about 800 on average per year, which suggests the importance of border effects for international economics.

We collect all estimates of the semi-elasticity from the primary studies. The approach yields an unbalanced data set, since some studies report many more estimates than other studies, but has three big advantages. First, it is demanding and sometimes impossible to select the authors' preferred estimate that would represent each study, so by collecting all estimates we avoid the most subjective stage of meta-analysis. Second, throwing away information is inefficient, and many studies report estimates employing alternative methods or data sets, which increases the variation in our data set. Third, using multiple estimates per study we can employ study-level fixed effects, which removes all characteristics idiosyncratic to individual studies. In total, we gather 1,271 estimates of the semi-elasticity; the median primary study reports 13 estimates.

A few problems concerning data collection are worth mentioning. To start with, the variable capturing the border effect is not always defined in the same way as

Figure 2.2: Estimated border effects vary widely



Notes: The figure shows a box plot of the estimates of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows) reported in individual studies. Full references for studies included in the meta-analysis are available in the online appendix.

Same country in (2.1). Often it equals one for cross-border trade flows, in which case we simply take the negative of the estimated coefficient. Sometimes, however, the dummy variable equals one only for trade flows crossing the border in one direction (for example, Anderson & Smith 1999). Following the common practice to "better err on the side of inclusion" in meta-analysis (Stanley 2001, p. 135), we choose to include the estimates of directional border effects, but control for this aspect of methodology to see whether it yields systematically different estimates. Finally, the collection of data is labor-intensive, since we gather information on 32 aspects of estimation design for all 1,271 estimates. To alleviate the danger of typos and mistakes, both of us collect the data independently and correct the inconsistencies by comparing the two data sets. The final data set is available in the online appendix at `meta-analysis.cz/border`.

Table 2.1: Border effects differ across countries

		Unweighted			Weighted		
	No. of estimates	Mean	95% conf. int.		Mean	95% conf. int.	
Canada	213	2.86	2.66	3.06	2.81	2.58	3.05
US	64	0.72	0.03	1.40	1.36	0.99	1.73
EU	263	2.55	2.04	3.05	2.59	2.18	2.99
OECD	98	2.35	1.71	3.00	2.41	1.90	2.91
Emerging	82	5.05	4.59	5.51	4.14	3.18	5.10
All countries	1,271	3.03	2.54	3.53	2.59	2.23	2.95

Notes: The table presents mean estimates of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows) for selected countries and country groups. Confidence intervals around the mean are constructed using standard errors clustered at both the study and data set level (the implementation of two-way clustering follows Cameron *et al.* 2011). In the right-hand part of the table estimates are weighted by the inverse of the number of estimates reported per study.

Figure 2.2 shows a box plot of estimates reported in the primary studies; the heterogeneity both between and within studies is substantial. It is apparent, however, that most studies report at least some estimates close to 3, near the original estimate provided by McCallum (1995). A large portion of heterogeneity in the estimates can be due to differences in data, and especially different countries for which the border effect is evaluated. Table 2.1 shows mean estimates of the countries and country groups that are examined most commonly in the literature.

We say that an estimate corresponds to the border effect of a particular country if identification of the semi-elasticity comes from trade flows within the country. For example, if data on trade flows between Canadian provinces are used, such as in McCallum (1995), we consider the estimated border effect Canadian, although the estimation also includes data on the US (flows between Canadian provinces and US states). Some authors used both province-to-province trade flows and state-to-state flows (for example, Anderson & van Wincoop 2003); the resulting estimates of the border effect correspond to both Canada and the US and are not shown in the table. The estimates for all other countries and groups of countries are nevertheless included in the overall mean reported in the last row of the table.

Table 2.1 also shows the corresponding confidence intervals constructed using clustered standard errors. Many meta-analyses cluster standard errors at the study level, because estimates reported in the same primary study are likely to be dependent. Nevertheless, we are not aware of any meta-analysis that would also try to take into account the dependence in estimates due to the use of similar data sets. A few studies in our sample use the same data set, especially the one introduced by Anderson & van Wincoop (2003), but many others simply add a few years to data used elsewhere. So we consider data sets to be the same or very similar if they provide data on the same region and start in the same year, and additionally cluster standard errors at the level of similar data sets. The implementation of two-level clustering follows the approach of Cameron *et al.* (2011).

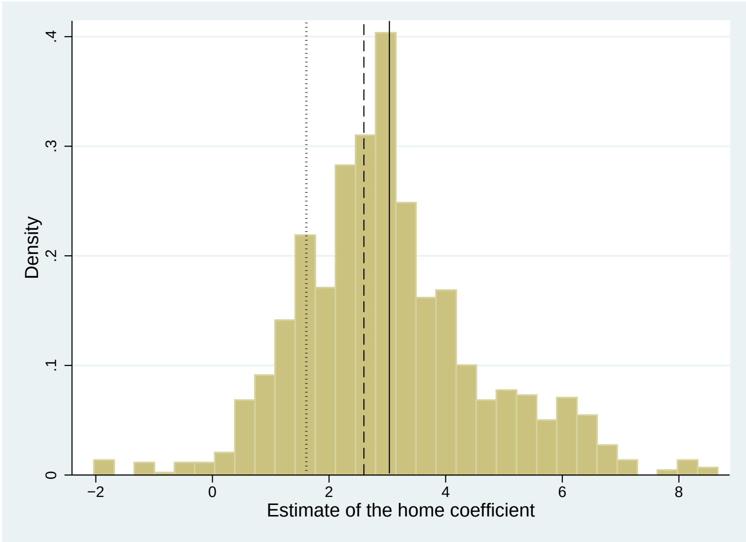
The left-hand part of the table shows unweighted estimates; the right-hand part shows estimates weighted by the inverse of the number of observations reported in each study. By using these weights we assign each study the same importance; otherwise studies reporting many semi-elasticities drive the results. The mean unweighted estimate of the semi-elasticity equals 3, virtually identical to the original estimate of the parameter by McCallum (1995). This semi-elasticity implies a border effect of $\exp(3) = 20$, which means that an average region in an average country trades twenty times more with regions in the same country than with foreign regions of similar characteristics. The 95% confidence interval for the mean estimate of the border effect is (13, 34), which shows substantial uncertainty due to differences in methodology.

The table documents that the semi-elasticities estimated for individual countries vary substantially. The smallest mean estimate corresponds to the US (implying a border effect of 2 in the case of the unweighted estimates), the largest mean is obtained for emerging countries (implying a border effect of 156). The respective means for Canada, EU, and OECD countries are close to the overall mean. When we weight the estimates by the inverse of the number of observations reported in each study, we obtain a smaller overall mean, implying a border effect of 13.3, and the country-specific estimates get less dispersed. In both cases is the lower bound of the 95% confidence interval of the estimate for emerging countries larger than the upper bounds of confidence intervals for all other groups of countries. That is, the border effects estimated in the literature suggest that developing and transition countries are substantially less integrated in global trade than developed countries.

Figure 2.3 shows the histogram of the estimated semi-elasticities. We see that almost all estimates are positive; in the data we only have 22 negative estimates, 1.7% of all semi-elasticities. The median estimate is very close to the overall mean and equals 2.9. The median estimate of median semi-elasticities reported in individual studies equals 2.6, which is virtually identical to the mean of estimates weighted by the inverse of the number of estimates reported per study. The closeness of the mean and median together with the shape of the histogram suggest that there are no serious outliers in our data set, so we do not exclude any estimates from the meta-analysis.

The journals in which primary studies are published differ greatly in prestige and rating. On the one hand, some studies are published in top field and general interest journals; on the other hand, many estimates come from studies published in local outlets. To illustrate the potential differences in quality we distinguish a group of studies published in top field or top or second-tier general interest journals: the *American Economic Review*, *Journal of International Economics*, *International Economic Review*, *European Economic Review*, and *Journal of Applied Econometrics*.

Figure 2.3: Studies in top journals report smaller estimates



Notes: The figure shows the histogram of the estimates of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows) reported in individual studies. The solid vertical line denotes the median of all estimates. The dashed line denotes the median of median estimates from studies. The dotted line denotes the median of estimates reported in studies published in the *American Economic Review*, *Journal of International Economics*, *International Economic Review*, *European Economic Review*, and *Journal of Applied Econometrics*.

Eleven studies in our sample are published in these journals and they report a median semi-elasticity of 1.7, implying a border effect of 5.5, less than a third of the overall mean effect. Studies in respected journals seem to report smaller semi-elasticities, but the pattern may be explained by differences in methodology. Another potential reason for between-study differences in estimates is publication selection.

2.3 Publication Bias

Publication selection bias arises when estimates have a different probability of being reported based on their sign or statistical significance. Sometimes it is called the "file drawer problem" (Rosenthal 1979): researchers may hide to their file drawers the estimates that are insignificant or have and unintuitive sign and search for estimates that are easier to publish. Publication bias has been identified in empirical economics by, for example, DeLong & Lang (1992), Card & Krueger (1995), and

Ashenfelter *et al.* (1999), among others. In a survey of examinations of publication bias, Doucouliagos & Stanley (2013) find that most fields of empirical economics are seriously affected by the problem. Because the potential presence of publication bias determines the weights that should be used in meta-analysis, we test for the bias before we proceed to the analysis of heterogeneity.

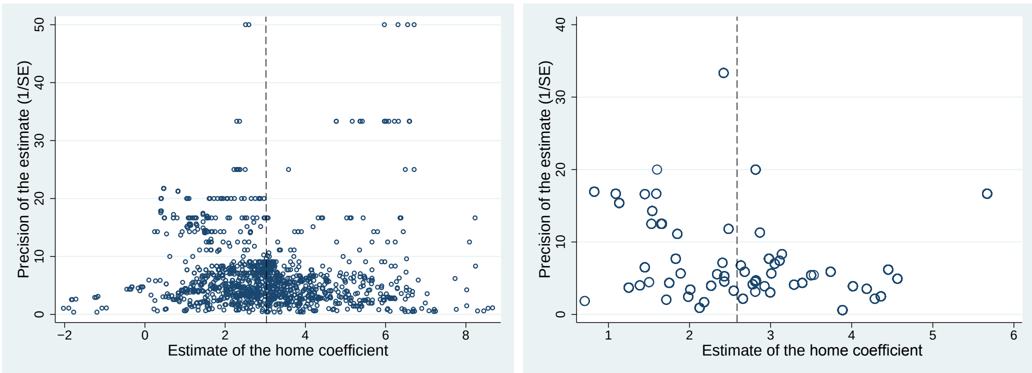
If researchers preferentially report estimates that are statistically significant and have the expected sign, the literature as a whole exaggerates the effect in question. For example, Stanley (2005) finds that the mean estimate of the price elasticity of water demand is exaggerated fourfold because of publication bias. The problem is widely recognized in medical science, and the best medical journals now require registration of clinical trials before publication, so that researchers can find the results of all trials, even though some are not submitted for publication. In a similar vein, the American Economic Association has agreed to establish a registry of randomized experiments "to counter publication bias" (Siegfried 2012, p. 648).

The presence of publication bias can be examined visually using the so-called funnel plot (Egger *et al.* 1997). It is a scatter plot showing the magnitude of estimated effects on the horizontal axis and the precision (the inverse of the estimated standard error) on the vertical axis. If the literature is not influenced by publication bias, the most precise estimates of the effect will be close to the mean underlying effect. As precision decreases, estimates get more dispersed, forming a symmetrical inverted funnel. In the presence of publication bias the funnel becomes asymmetrical (if researchers discard estimates of a particular sign or magnitude), or hollow (if researchers discard statistically insignificant estimates), or both.

We report the funnel plot for the border effect literature in Figure 2.4. Panel (a) shows the funnel for all estimates; panel (b) only shows median estimates for each study. We make three observations from the funnels. First, both funnels are relatively symmetrical, with the most precise estimates being close to the average reported semi-elasticity. Second, the funnels are not hollow, and even estimates with very little precision (and, thus, small p-values) are reported. Three, the funnel in panel (a) has multiple peaks, which suggests heterogeneity in the estimated border effects. Signs of heterogeneity are not surprising given our estimates of cross-country differences in the previous section. We conclude that typical funnel plots reported in economics meta-analyses show much clearer signs of publication bias than what we observe in the literature on border effects (see, for example, Stanley & Doucouliagos 2010).

Figure 2.4: Funnel plots suggest little publication bias

(a) All estimates (b) Median estimates reported in studies



Notes: In the absence of publication bias the funnel should be symmetrical around the most precise estimates of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows). The dashed vertical lines denote the mean of all estimates in panel (a) and the mean of median estimates reported in studies in panel (b). Multiple peaks of the funnel suggest heterogeneity.

The funnel plot represents a simple visual tool for the evaluation of publication bias, but the presence of bias can be tested more formally. Following Card & Krueger (1995), we explore the relationship between estimates of the semi-elasticity and their standard errors. Because researchers estimating the semi-elasticity assume that the estimates have a t-distribution, the reported semi-elasticities should be distributed approximately normally around the mean reported effect. In contrast, if statistically significant estimates are preferred, researchers will search for large estimates of the semi-elasticity in order to offset the standard errors and produce large t-statistics. Similarly, when researchers discard negative estimates, a positive relationship arises between the reported estimates and their standard errors because of heteroskedasticity (Stanley 2008):

$$HOME_{ij} = HOME_0 + \beta \cdot SE(HOME_{ij}) + u_{ij}, \tag{2.2}$$

where $HOME_{ij}$ are estimates of the semi-elasticity, $SE(HOME_{ij})$ are the reported standard errors of the semi-elasticity estimates, $HOME_0$ is the mean semi-elasticity corrected for potential publication bias, β measures the extent of publication bias, and u_{ij} is a normal disturbance term. For example, if the true mean semi-elasticity was zero (implying no border effect) but all researchers reported the 5% of estimates that are positive and statistically significant, the estimated $HOME_0$ would be close to two: the researchers would need their t-statistics, $HOME/SE(HOME)$, to equal at least two.

Equation (2.2) can be interpreted as a test of funnel asymmetry, because it follows from rotating the axes of the funnel plot and inverting the values on the new horizontal axis to show standard errors instead of precision. Note that the test has low power if the true underlying effect is close to zero and the only source of publication bias is selection for statistical significance: when $HOME_0$ is zero and insignificant estimates, positive or negative, are omitted, β is zero, even though publication selection may be substantial (the funnel plot gets hollow, but not asymmetrical). Nevertheless, such a symmetrical selection does not create a bias in the mean of the reported estimates, so it is usually not a source of concern (Stanley 2005).

We present the results of funnel asymmetry tests in Table 2.2. Because (2.2) is heteroskedastic, we present robust standard errors, which are clustered at the level of individual studies and data sets. The first column of panel A shows estimates of the parameters from (2.2) using all 1,271 semi-elasticities in our sample. The coefficient corresponding to the extent of publication bias is statistically insignificant and close to zero, while the estimated semi-elasticity beyond publication bias is 2.9, close to the mean and median semi-elasticity reported in the literature. Therefore neither visual nor formal tests show any evidence of publication selection, and the potential selection does not create any bias in the mean reported estimate of the border effect.

Table 2.2: Funnel asymmetry tests show no publication bias

Panel A: unweighted regressions	All estimates	Published	Fixed effects	Instrument
SE (publication bias)	0.604	0.599	0.383	-0.797
	(0.514)	(0.522)	(0.534)	(2.020)
Constant (effect beyond bias)	2.852***	2.932***	2.918***	3.270***
	(0.321)	(0.339)	(0.159)	(0.724)
Studies	61	48	61	61
Observations	1,271	1,144	1,271	1,271
Panel B: weighted regressions	Precision	Study	Impact	Citations
SE (publication bias)	0.246	1.489	3.062	5.073
	(1.964)	(1.170)	(2.024)	(4.272)
Constant (effect beyond bias)	2.959***	2.204***	1.634***	1.235**
	(0.723)	(0.395)	(0.424)	(0.501)
Studies	61	61	53	49
Observations	1,271	1,271	1,124	1,069

Notes: The table presents the results of regression $HOME_{ij} = HOME_0 + \beta \cdot SE(HOME_{ij}) + u_{ij}$. $HOME_{ij}$ and $SE(HOME_{ij})$ are the i-th estimates of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows) and their standard errors reported in the j-th studies. Standard errors of regression parameters are clustered at both the study and data set level and shown in parentheses (the implementation of two-way clustering follows Cameron *et al.* 2011). Published = we only include published studies. Fixed effects = we use study dummies. Instrument = we use the number of observations in the gravity equation as an instrument for the standard error. Regressions in Panel B are estimated by weighted least squares. Precision = we take the inverse of the reported estimate's standard error as the weight. Study = in addition to "Precision" the inverse of the number of estimates reported per study is taken as the weight. Impact = in addition to "Study" the RePEc resursive discounted impact factor of the outlet where the study was published is taken as the weight. Citations = in addition to "Impact" the number of Google Scholar citations received per year is taken as the weight. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

The second column of panel A in Table 2.2 estimates equation (2.2) using only the semi-elasticities reported in published studies. Perhaps editors or referees prefer large and statistically significant coefficients, which would pull the mean reported semi-elasticity up. Indeed, in a meta-analysis of vertical productivity spillovers from foreign direct investment, Havranek & Irsova (2011) find that studies published in refereed journals show substantially more publication bias than unpublished manuscripts. Our results concerning the border effect, however, show little difference between published and unpublished studies both in the extent of publication bias and the mean underlying semi-elasticity beyond any potential bias.

Next, we include fixed effects for individual studies to control for method or other quality characteristics specific to individual studies. The fixed-effects estimation represents another advantage of collecting multiple estimates per study. The results are very similar to the baseline specification reported in the first column; we get no evidence of publication bias, and the mean estimated semi-elasticity is still 2.9.

Specification (2.2) only includes one explanatory variable, the standard error. It is possible that some method choices affect both the estimated semi-elasticity and the corresponding standard error, which would cause the error term u_{ij} to be correlated with $SE(HOME_{ij})$. In the last column of panel A in Table 2.2 we use the logarithm of the number of observations in the gravity equation as an instrument for $SE(HOME_{ij})$: the number of observations is correlated with the reported standard errors of the semi-elasticities, but little related to the methods of estimation. The instrumental variable estimation is less precise, but still reports the mean underlying semi-elasticity close to 3 and no evidence of publication bias.

In panel B of Table 2.2 we weight all estimates by precision. We have noted that equation (2.2) is heteroskedastic, and the explanatory variable directly captures the variance of the response variable. To achieve efficiency, many applications of meta-analysis divide (2.2) by the corresponding standard error; that is, multiply the equation by the precision of the estimates. Such an approach has the additional allure of giving more importance to precise results. The first column of panel B shows that weighting by precision has little impact on our results.

The second column of panel B adds weighting by the inverse of the number of estimates reported in studies to precision weights. In line with the summary statistics from the previous section, the mean semi-elasticity decreases when each study gets the same weight. Next, in column 3 we add weighting by the discounted recursive RePEc impact factor of the publication outlet. The estimated semi-elasticity decreases to 1.6: better journals seem to publish smaller estimates, which corroborates our interpretation of Figure 2.3. Finally, we also weight the estimates by the the number of Google Scholar citations the study receives each year. The semi-elasticity decreases to 1.3, implying a border effect of 3.4. Thus, when we give more weight to highly-cited papers published in good journals, we are able to shrink the mean border effect more than five times. In the next section we explore how these differences between studies can be explained by variation in data and methodology.

2.4 Why Border Effects Vary

2.4.1 Variables and Estimation

We substitute the characteristics of estimates and studies for $SE(HOME_{ij})$ in equation (2.2). The previous section shows that the reported standard errors are not correlated with the estimates of the semi-elasticity, and the exclusion of the standard error has the additional benefit of removing the obvious heteroskedasticity. After we remove the standard error from the equation, we have little to gain by weighting our estimates by precision. Moreover, weighting by the estimates' precision introduces artificial variation to variables defined at the study level (for example, the use of disaggregated or panel data). Instead we weight regressions by the inverse of the number of estimates reported per study to give each study the same weight, and also report a robustness check using unweighted data.

Table 2.3 lists all the variables that we collect from primary studies, explains their definition, and shows summary statistics. The last column presents the mean weighted by the inverse of the number of estimates reported in each study. We divide the variables into seven groups:

First, we collect information on data characteristics. Second, we control for regional differences in the estimates. Three, we collect variables reflecting the general design of the analysis. Four, we include dummy variables that capture how the authors treat multilateral resistance. Five, we distinguish between the different types of the treatment of zero trade flows. Six, we include dummy variables reflecting whether the gravity equation uses control variables. Finally, we include information on publication and citation characteristics of the studies. Our intention is to introduce the possible reasons for heterogeneity in the estimated border effects, not to present a detailed survey of the methods used in estimating the gravity equation. For a survey of methods see Head & Mayer (2014).

Table 2.3: Description and summary statistics of regression variables

Variable	Description	Mean	SD	WM
Home	The coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows (or minus the coefficient on the dummy variable that equals one for cross-border flows).	3.03	1.60	2.59
SE	The estimated standard error of *home*.	0.30	0.35	0.26
Data characteristics				
Midyear of data	The midpoint of the sample on which the gravity equation is estimated (base: 1899).	91.3	16.0	91.7
Panel data	= 1 if panel data are used in the gravity equation.	0.67	0.47	0.52
Disaggregated	= 1 if trade flows are disaggregated at the sector or product level.	0.57	0.50	0.41
Obs. per year	The logarithm of the number of observations per year included in the gravity equation.	6.89	1.31	6.93
No. of years	The logarithm of the number of years in the data.	1.27	1.04	0.91
Countries examined				
Canada	=1 if the border effect is estimated for Canada.	0.17	0.37	0.18
US	=1 if the border effect is estimated for the US.	0.05	0.22	0.08
EU	=1 if the border effect is estimated for the EU.	0.21	0.41	0.23
OECD	=1 if the border effect is estimated for OECD countries.	0.08	0.27	0.06
Emerging	=1 if the border effect is estimated for developing or transition countries.	0.06	0.25	0.05
Design of the analysis				
No internal trade	=1 if within-country trade flows are not observed but estimated using production data.	0.58	0.49	0.43
Inconsistent dist.	=1 if within-country distance is measured differently from between-country distance.	0.14	0.35	0.21
Actual distance	=1 if actual distance traveled by road or sea is used instead of the great-circle formula.	0.06	0.24	0.07
Total trade	=1 if total trade is used as the dependent variable and imports and exports are summed before taking logs.	0.01	0.12	0.01
Asymmetry	=1 if the estimate measures the difficulty of cross-border flows in one direction.	0.29	0.45	0.14
Instruments	=1 if instruments are used to correct for the endogeneity of GDP.	0.06	0.25	0.06
Treatment of multilateral resistance				
Remoteness	=1 if remoteness terms are included.	0.06	0.24	0.10
Country fixed eff.	=1 if destination and origin fixed effects are included.	0.27	0.44	0.31
Ratio estimation	=1 if trade flows are normalized by trade with self.	0.31	0.46	0.11
Anderson est.	=1 if the nonlinear estimation method developed by Anderson & van Wincoop (2003) is used.	0.02	0.15	0.06
No control for MR	=1 if the gravity equation does not account for multilateral resistance terms.	0.38	0.49	0.50
Treatment of zero trade flows				
Zero plus one	=1 if one is added to observations of zero trade flows.	0.11	0.32	0.13
Tobit	=1 if the gravity equation is estimated by the Tobit model.	0.06	0.24	0.06
PPML	=1 if the gravity equation is estimated by the Poisson pseudo-maximum likelihood estimator.	0.07	0.26	0.11
Zeros omitted	=1 if observations of zero trade flows are omitted.	0.66	0.47	0.55
Control variables				
Adjacency cont.	= 1 if the gravity equation controls for adjacency.	0.63	0.48	0.50

Continued on next page

Description and summary statistics of regression variables (continued)

Variable	Description	Mean	SD	WM
Language control	= 1 if the gravity equation controls for shared language (when needed).	0.78	0.42	0.73
FTA control	= 1 if the gravity equation controls for free trade agreements (when needed).	0.73	0.44	0.76
Publication characteristics				
Published	= 1 if the study is published in a peer-reviewed journal.	0.90	0.30	0.79
Impact	The recursive discounted RePEc impact factor of the outlet (collected in January 2014).	0.46	0.90	0.45
Citations	The logarithm of the mean number of Google Scholar citations received per year since the study appeared in Google Scholar (collected in January 2014).	1.52	1.13	1.60
Publication year	The year when the study first appeared in Google Scholar (base: 1995).	9.46	4.32	9.62

Notes: SD = standard deviation. WM = mean weighted by the inverse of the number of estimates reported per study. All variables except for citations and the impact factor are collected from studies estimating the border effect (the search for studies was terminated on January 1, 2014, and the list of studies is available in the online appendix. Citations are collected from Google Scholar; the impact factor from RePEc. The data set is available in the online appendix at `meta-analysis.cz/border`.

Data characteristics We control for the age of the data by creating a variable that reflects the midpoint of the sample; perhaps the mean border effect shrinks with the continuing globalization and integration of emerging markets. The mean semi-elasticity in our sample is estimated using data from 1990. To see whether cross-sectional and panel data yield systematically different border effects, we include a corresponding dummy variable. Sixty-seven per cent of estimates come from specifications using panel data, but 48% of studies rely on cross-sectional data (that is, panel studies usually report more estimates).

Next, we control for the level of aggregation in the gravity equation and add a dummy that equals one if the data are disaggregated at the sector or product level; about a half of all studies employ some sort of disaggregation. Researchers suspect that aggregation across products and sectors creates a bias in the gravity equation, but the direction of the bias is unclear (Anderson & van Wincoop 2004, pp. 727-729). We also include the logarithm of the number of observations per year used in the gravity equation and the logarithm of the number of years in the panel. The mean semi-elasticity in our sample is computed using 3 years of data and 1,000 estimates per year.

Countries examined Border effects in our sample are estimated for different regions, so we control for regional differences. Among other things, countries may display different elasticities of substitution between domestic and foreign goods, which would affect the estimated border effect. We include five regional dummies: Canada, the US, EU, OECD, and emerging countries (including both developing and transition economies). The first paper on the border effect, McCallum (1995), uses data on internal trade in Canada. Many others have followed, and 17% of all estimates in our sample use Canadian data. Another 5% of border effects are estimated for the US (for example, Anderson & van Wincoop 2003), 21% for the EU (Nitsch 2000), 8% for the OECD (Wei 1996), and 6% for emerging countries (da Silva *et al.* 2007). The remaining reported elasticities are estimated for other individual OECD countries or use combinations of internal trade flows for different regions.

Design of the analysis We distinguish studies that have data on within-country trade flows from studies that estimate trade with self using production data; about a half of the studies have access to data on internal trade. Regarding the studies that must compute data on trade with self, we distinguish between those that use the same definition for the computation of within and between-country distance and those that employ different definitions. Head & Mayer (2010) show that employing inconsistent measures of internal distance can exaggerate the reported border effect. About 14% of all estimates are obtained using different definitions of internal and external distance.

We also include a dummy variable that equals one for estimates obtained with a measure of distance computed from actual road or sea routes instead of the great-circle formula (6% of all estimates). We expect that the great-circle formula overstates internal distance and thus leads to an upward bias in the estimated border effect. Regions are likely to be connected more efficiently with other regions in the same country than with foreign regions that show the same great-circle distance (Braconier & Pisu 2013). A couple of studies in our data set commit what Baldwin & Taglioni (2007) call the "silver medal mistake" of estimating the gravity equation:

they use total or average trade flows as the response variable and compute the sum or average before taking logs. About 14% of studies use an asymmetric definition of border effects, which means that they examine the difficulty of crossing borders in one direction (for example, Anderson & Smith 1999). Finally, we control for the case when researchers use instruments to account for the endogeneity of GDP in the gravity equation (6% of all estimates).

Treatment of multilateral resistance We include five dummy variables to control for the way the authors of primary studies account for the problem. The first attempts, usually prior to Anderson & van Wincoop (2003), involve including remoteness terms, and about 10% of studies in our sample do so. The most straightforward approach is to use destination and origin fixed effects (Feenstra 2002), employed by 31% of studies. Another consistent estimation method involves normalizing trade flows by trade with self (Head & Mayer 2000), and 11% of studies use this method. About 6% of studies use the non-linear technique introduced by Anderson & van Wincoop (2003). A half of the primary studies do not estimate the border effect consistently; that is, they either add the atheoretical remoteness terms or ignore multilateral resistance entirely.

Treatment of zero trade flows The simplest way to incorporate zeros is to add one to each observation and use the log-linear transformation. But as Head & Mayer (2014) note, in this case the results depend on the units of measurement. Many authors who choose this approach estimate the gravity equation using Tobit (6% of the studies). Next, 11% of primary studies use the non-linear method introduced by Silva & Tenreyro (2006), Poisson pseudo-maximum likelihood estimator (PPML). The method allows for the incorporation of zero trade flows and addresses heteroskedasticity in the error term of the gravity equation. Finally, 55% of studies exclude zeros from their data sets. Some studies, especially those using aggregated OECD data, do not face the problem because they have no zero trade flows in their data.

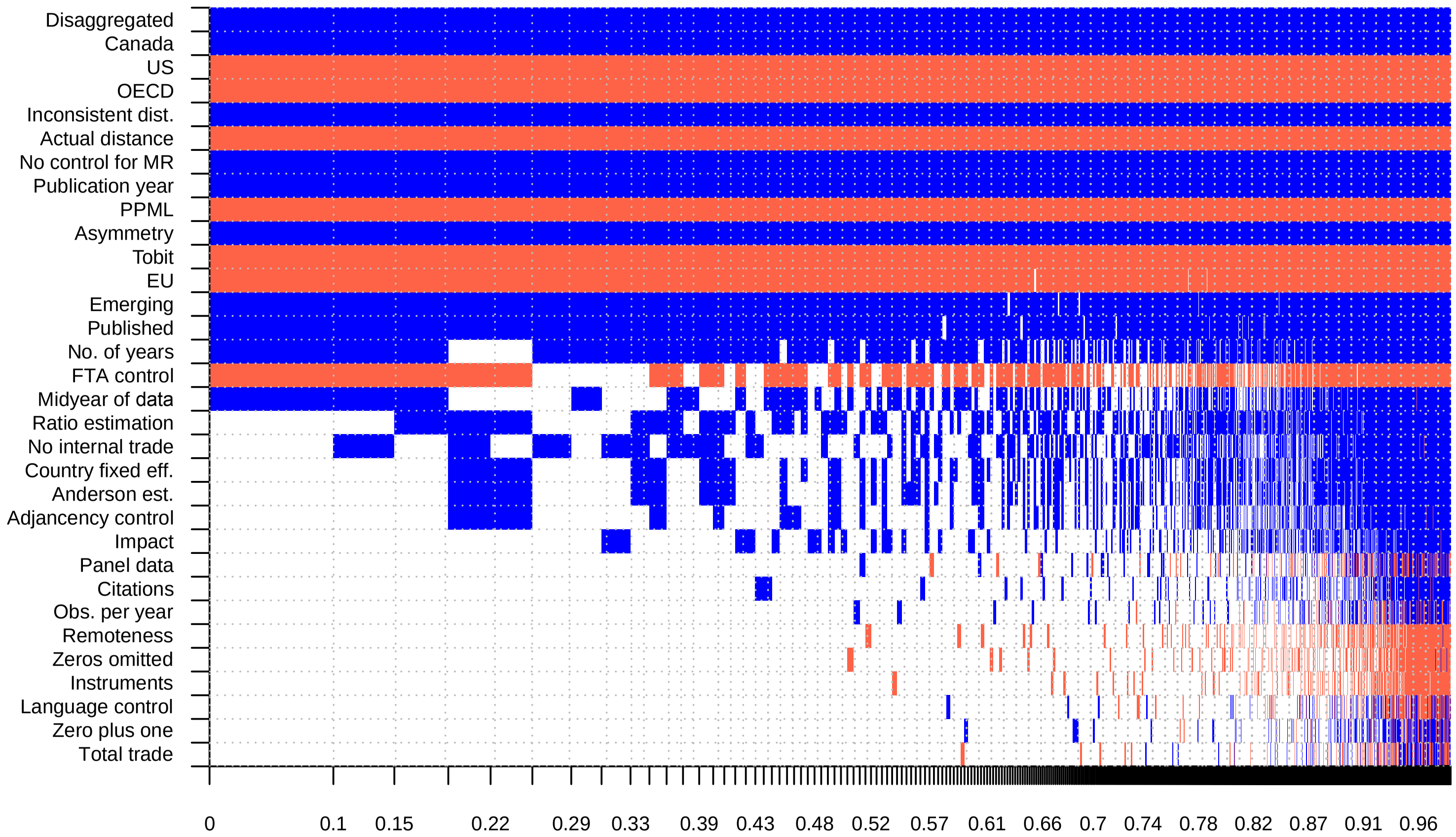
Control variables Studies estimating the border effect typically include three control variables: a dummy for adjacency, common language, and membership in a free trade agreement. We examine whether the inclusion of these variables has a systematic influence on the estimated semi-elasticity. In many cases the primary studies cannot include the dummy variables for common language and free trade area membership, because the value of these dummies would be the same for all trading pairs in their data—for example, trade flows between Canadian provinces and US states. We code the variables in the way that "0" for common language and FTA control means that the control variable could be included but is omitted.

Publication characteristics To see whether published studies yield different results even after all the main aspects of methodology are controlled for, we include a dummy variable that equals one if the study is published in a peer-reviewed journal. To account for the different quality of publication outlets, we include the recursive discounted RePEc impact factor. The greatest advantage of RePEc with respect to other impact metrics is that it provides information on virtually all journals and working paper series. Next, we control for the number of citations of the study, which could reflect aspects of study quality not captured by the data and methodology variables described above. Finally, for each study we find the year when it first appeared in Google Scholar and examine whether there is a publication trend in the estimates of the border effect beyond advances in methodology.

We intend to run a regression with the semi-elasticity as the response variable and all the aspects of data, methodology, and publication as explanatory variables. The problem is that such a regression would likely contain many redundant variables, and we do not know a priori which of the variables introduced in Table 2.3 should be excluded. Ideally we would also like to run regressions containing different subsets of the explanatory variables to see whether our results are robust. We face model uncertainty, which can be addressed by Bayesian model averaging (BMA).

BMA runs many regressions involving different subsets of the 32 potential explanatory variables. With 2^{32} possible combinations, it would take several months

Figure 2.5: Model inclusion in BMA



Notes: Response variable: estimate of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows). All regressions are weighted by the inverse of the number of estimates reported per study. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures cumulative posterior model probabilities. Numerical results of the BMA estimation are reported in Table 2.4.

to estimate all the regressions, so our approach relies on a Monte Carlo Markov Chain algorithm that walks through the potential models (we use the `bms` R package by Feldkircher & Zeugner 2009). For each model BMA computes a weight, called posterior model probability, which is analogous to information criteria or adjusted R-squared and captures how well the model fits the data. The regression coefficients reported by BMA are weighted averages of the many estimated models; instead of standard errors, BMA reports posterior standard deviations reflecting the distribution of regression parameters retrieved from individual models. For each variable we compute posterior inclusion probability, which is the sum of posterior model probabilities of the regressions in which the variable is included. Posterior inclusion probability reflects how likely it is that the variable should be included in the true model. Diagnostics of our BMA exercise is available in Appendix A). More details on BMA in general can be found, for example, in Raftery *et al.* (1997) or Eicher *et al.* (2011).

2.4.2 Results

Figure 2.5 reports our results concerning model inclusion of different explanatory variables in the BMA exercise. The columns in the figure show different regression models, and the width of the columns denotes posterior model probability. The rows show individual variables, sorted by posterior inclusion probability in a descending order. If the cell corresponding to the variable is empty, it means that the variable is not included in the model. Blue color (darker in grayscale) means that the variable is included and the estimated sign of the regression parameter is positive. Red color (lighter in grayscale) denotes a negative estimated regression parameter. We can see that approximately a half of the variables appear in the best models and that the signs of their estimated regression parameters are robust to including other control variables.

The numerical results of Bayesian model averaging are reported in Table 2.4. In addition, we show results of an OLS regression which includes all but the 11 variables with posterior inclusion probability lower than 0.3: these 11 variables do not help explain the variability in the estimates of the border effect. The OLS estimation produces results consistent with those of BMA. The estimated signs of regression parameters are the same and variables with high posterior inclusion probability in BMA are usually statistically significant in the OLS estimation. Also the estimated magnitudes of regression parameters are similar in both methods for the most important variables; that is, those with high posterior inclusion probabilities. When interpreting posterior inclusion probability, we follow the approach of Eicher *et al.* (2011), who considers the value as *weak* if it is between 0.5 and 0.75, *substantial* if it is between 0.75 and 0.95, strong if it is between 0.95 and 0.99), and *decisive* it the posterior inclusion probability exceeds 0.99.

Some of the data characteristics systematically affect the reported estimates of the border effect. Researchers using disaggregated data tend to obtain estimates of the semi-elasticity larger by 0.8; the posterior inclusion probability of this variable is decisive. The result corroborates the findings of Anderson & Yotov (2010, p. 2167), who also report that aggregated data yield smaller estimates of the border effect. In contrast, Hillberry (2002) finds that aggregation exaggerates the border effect. Next, more years of the data available for estimation translates into larger border effects, but the posterior inclusion probability of this variable is only 0.81. For all other variables in this category we get weak posterior inclusion probabilities.

Regional differences help to explain the heterogeneity in the estimated border effects; the posterior inclusion probabilities for all the region dummies are decisive. Researchers typically obtain the largest border effects for developing and transition countries, followed by Canada. The smallest estimates are reported for the US. Balistreri & Hillberry (2007) discuss how the small estimates for the US can be affected by the characteristics of the Commodity Flow Survey, the source of the data typically used for this estimation.

Regarding the general design of the gravity equation, it matters for the estimated border effect whether internal and external distances are measured consistently. If not, the reported semi-elasticities tend to be larger by about 0.8; the result is in line with the findings of Head & Mayer (2010). When the authors of primary studies use

Table 2.4: Explaining the differences in the estimates of the border effect

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of Home	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value
Data characteristics						
Midyear of data	0.003	0.004	0.542	0.001	0.011	0.915
Panel data	0.004	0.055	0.068			
Disaggregated	0.800	0.138	1.000	0.654	0.359	0.069
Obs. per year	0.001	0.008	0.048			
No. of years	0.136	0.079	0.811	0.147	0.107	0.170
Countries examined						
Canada	0.718	0.126	1.000	0.741	0.322	0.021
US	-1.177	0.134	1.000	-1.135	0.239	0.000
EU	-0.518	0.165	0.992	-0.639	0.391	0.102
OECD	-0.981	0.176	1.000	-0.958	0.356	0.007
Emerging	0.947	0.267	0.990	0.808	0.388	0.037
Design of the analysis						
No internal trade	0.166	0.210	0.441	0.491	0.404	0.224
Inconsistent dist.	0.783	0.142	1.000	0.514	0.302	0.089
Actual distance	-0.933	0.153	1.000	-0.666	0.313	0.033
Total trade	0.000	0.049	0.025			
Asymmetry	0.536	0.121	0.999	0.540	0.246	0.028
Instruments	-0.005	0.043	0.035			
Treatment of multilateral resistance						
Remoteness	-0.007	0.045	0.048			
Country fixed eff.	0.213	0.311	0.368	0.220	0.305	0.471
Ratio estimation	0.402	0.475	0.520	0.602	0.584	0.303
Anderson est.	0.229	0.347	0.350	0.079	0.353	0.822
No control for MR	0.826	0.299	1.000	0.719	0.308	0.019
Treatment of zero trade flows						
Zero plus one	0.001	0.023	0.029			
Tobit	-0.636	0.156	0.996	-0.553	0.312	0.077
PPML	-0.707	0.154	1.000	-0.774	0.493	0.117
Zeros omitted	-0.004	0.026	0.042			
Control variables						
Adjancency control	0.071	0.136	0.258			
Language control	-0.001	0.018	0.030			
FTA control	-0.213	0.177	0.661	-0.366	0.347	0.292
Publication characteristics						
Published	0.339	0.108	0.976	0.330	0.265	0.212
Impact	0.018	0.044	0.183			
Citations	0.003	0.014	0.063			
Publication year	0.075	0.012	1.000	0.058	0.031	0.062
Constant	0.087	NA	1.000	0.922	1.058	0.383
Studies	61			61		
Observations	1,271			1,271		

Notes: Home = the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows. PIP = posterior inclusion probability. SD = standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.3. Standard errors in the frequentist check are clustered at both the study and data set level (the implementation of two-way clustering follows Cameron *et al.* 2011). More details on the BMA estimation are available in Table 2.8 and Figure 2.6. A detailed description of all variables is available in Table 2.3.

actual road or sea distances instead of employing the great-circle formula, they report much smaller estimates of the semi-elasticity: by 0.9. Braconier & Pisu (2013) also find that using actual distance reduces the estimated border effect. Next, asymmetric estimates of the border effect are on average larger than the ones using the symmetric definition. The border effects estimated using data on trade with self computed from production statistics differ little from the estimates obtained when data on within-country trade are directly available. It seems that the "silver medal mistake" in estimation does not affect the resulting border effects, but very few papers in our data set commit the mistake.

In contrast, the "gold medal mistake" of estimating gravity equations has important consequences for the border effect: if authors do not control for multilateral resistance terms, they are likely to report semi-elasticities larger by 0.8. This result contrasts with the findings of Balistreri & Hillberry (2007), who report that the decrease in border effects found by Anderson & van Wincoop (2003) is primarily due to the specifics of the data and not due to the control for multilateral resistance. The posterior inclusion probabilities for the specific types of control for multilateral resistance are weak: when estimating the border effect, it is important to control for multilateral resistance, but it does not seem to matter how exactly it is done.

The treatment of zero trade flows affects the estimated border effect as well. If Tobit or PPML are used, the resulting semi-elasticities are smaller by about 0.7. In contrast, the inclusion of control variables does not seem to matter much for border effects. Concerning publication and other study characteristics, papers published in refereed journals tend to report semi-elasticities larger by about 0.3. The impact factor of the journal and the number of citations are not important for the reported border effects when we control for the characteristics of data and methods. The reported border effects seem to increase slightly in time: the semi-elasticities are on average larger by 0.075 each year.

In the next step we try to piece the puzzle together by computing a mean estimate of the border effect conditional on avoiding the gold medal, silver medal, or any other potential mistake in estimation. This part of our analysis is the most subjective, because it involves defining "best practice" in the estimation of border effects, and different researchers may have different opinions on what the best practice is. Nevertheless, we show that the major innovations introduced to the estimation of gravity equations in the last decade substantially alleviate the border puzzle, and seem to solve it at least for some regions.

For each variable in Table 2.4 we select a preferred value, or a sample mean if we have no preference, and compute the implied semi-elasticity as a linear combination of all the regression parameters. In other words, we construct a synthetic study with a large number of observations, best practice methodology, and maximum number of citations and other publication characteristics. We select sample maxima for the midyear of data (that is, we put emphasis on studies using recent data), panel data, disaggregated data, the number of observations per year, the number of years in the data, actual distance, PPML, inclusion of the control variables, publication in a refereed journal, the impact factor, and the number of citations. We plug in sample minima for the dummy variable corresponding to unavailability of within-country data, inconsistent measure of internal and external distance, summing trade flows before taking logs, estimating an asymmetric border effect, adding remoteness terms, disregarding multilateral resistance, adding one to zero trade flows and using Tobit for estimation, and disregarding zero trade flows. All other variables are set to their sample means.

Table 2.5 presents the results; the overall mean semi-elasticity is reported in the last row and region-specific estimates are in the remaining rows. The column labeled "Diff." shows the difference between our new estimates and the simple means reported in Table 2.1. The left-hand part of the table shows baseline results constructed from Table 2.4; the right-hand part is based on regressions not weighted by the inverse of the number of estimates reported per study (Table 2.7). Both sets of results are qualitatively similar, but the unweighted specification yields smaller estimates for all regions except Canada, and even reports negative semi-elasticities for the US and EU. We focus on the results obtained from the weighted regressions.

From Table 2.5 we see that giving more weight to studies that correct for the

Table 2.5: Advances in methodology shrink the border effect

	Weighted				Unweighted			
Best practice	Estimate	95% conf. int.		Diff.	Estimate	95% conf. int.		Diff.
Canada	1.95	1.09	2.81	-0.86	2.14	0.80	3.49	-0.72
US	0.06	-1.02	1.13	-1.30	-0.51	-1.73	0.71	-1.23
EU	0.72	-0.62	2.05	-1.87	-0.17	-1.60	1.25	-2.72
OECD	0.25	-1.12	1.62	-2.16	0.08	-1.40	1.55	-2.27
Emerging	2.18	0.67	3.69	-1.96	2.02	0.62	3.41	-3.03
All countries	1.13	0.04	2.23	-1.46	0.93	-0.43	2.29	-2.10

Notes: The table presents estimates of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows) for selected countries and country groups implied by Bayesian model averaging and our definition of best practice. That is, we take the regression coefficients estimated by BMA and construct fitted values of *home* conditional on control for multilateral resistance, consistent measure of within and between-country distance, and other aspects of methods and data (see text for details). Diff. = the difference between these estimates and the simple means reported in Table 2.1. Confidence intervals are approximate and constructed using the standard errors estimated by OLS. The right-hand part of the table presents results based on the robustness check using unweighted regressions (Table 2.7).

traditional problems in gravity equations and use novel methods decreases the estimated semi-elasticities significantly for each region. (The difference would be even larger if we plugged in sample means for publication characteristics and the number of observations and years in the data instead of giving more weight to large, broadly cited studies published in good journals.) The overall mean semi-elasticity is 1.1, which translates into a border effect of 3.1—almost seven times less than the border effect based on the sample mean of semi-elasticities reported in the literature. The border effect for the US and OECD countries is negligible, only $\exp(0.06) = 1.06$ and $\exp(0.25) = 1.28$; in contrast, the effect is still substantial for emerging countries: $\exp(2.18) = 8.85$. Regions in emerging countries tend to trade almost nine times more with regions in the same country than with similar foreign regions.

To put these numbers into perspective, we compute the ad-valorem tariff equivalent of the border effect. The tariff equivalent can be expressed as the following: $\exp(\text{home}/\text{trade costs elasticity}) - 1$, so we need an estimate of the elasticity of trade with respect to trade costs. We use the survey of Head & Mayer (2014), who find a median elasticity of 5.03 estimated in studies controlling for multilateral resistance and using tariff variation to identify the elasticity. For an average region the tariff equivalent is $\exp(1.13/5.03) - 1 = 25\%$. For OECD countries the tariff equivalent of border barriers falls to 5.2%, which is less than a half of the mean tariff equivalent of core non-tariff barriers to trade estimated by Kee *et al.* (2009), 12%. In contrast, our estimates of the border effect for emerging countries suggest a high tariff equivalent of 54%.

One of the main points of Anderson & van Wincoop (2003) is that the general equilibrium trade impact of borders, which takes into account price index, wage, and GDP changes in response to changes in trade costs, is smaller than the partial equilibrium impact reflected in the coefficient estimated in the gravity equation. We approximate the general equilibrium effect using our estimate of the partial equilibrium effect and the approach based on exact hat algebra (Dekle *et al.* 2007) described in Head & Mayer (2014, pp. 167-170, who also provide a Stata code for the computation). Employing the data provided by Head & Mayer (2014), bilateral trade flows of 84 countries for which values of internal trade can be computed, we obtain a general equilibrium border effect of 2.15 for regions in the same country and 0.72 for regions across borders. That is, our results suggest that for an average country borders reduce international trade by 28% and increase within-country trade by 115%.

2.5 Robustness Checks

We present two additional sets of results. First, we use alternative priors for Bayesian model averaging. Second, we employ unweighted regressions. We show that the results are similar to the baseline in terms of the estimated effects of the different aspects of study design on the estimated semi-elasticities, and that the resulting "best practice" estimates of the border effect are close to those reported in the previous section.

In the baseline specification we use the unit information prior for Zellner's g-prior, which means that the prior (each regression coefficient equals zero) provides the same amount of information as one observation in the data set. Because we have 1,271 observations, the prior does not drive the posterior results. The second important choice is the model prior, which determines the prior probability of each model. In the baseline specification we employ the uniform model prior, which gives each model the same prior probability. Eicher *et al.* (2011) shows that these intuitive priors yield the best predictive performance. Nevertheless, there are obviously many other ways how to choose the priors and the choice could influence our results.

The disadvantage of the uniform model prior is that it gives more weight to models with the mean number of variables, which is $32/2 = 16$ in our case. Such models appear most frequently among the subsets of all the 2^{32} possible models. Nevertheless, the true model may only contain a few variables, so the emphasis on large models can be contraproductive. An alternative is the beta-binomial prior advocated by Ley & Steel (2009), which gives the same prior probability to each *model size*, and thus does not prefer large models. An often-used alternative to the unit information prior is the BRIC g-prior (for example, Fernandez *et al.* 2001).

Table 2.6 summarizes the results of Bayesian model averaging with the alternative priors; we provide more details and diagnostics in Table 2.9 and Figure 2.7 in Appendix A. The results are very similar to our baseline specification concerning the estimated posterior inclusion probabilities for the explanatory variables, the signs of regression coefficients, and their magnitude. The semi-elasticity conditional on best practice is 1.02, implying a partial equilibrium border effect of 2.8, slightly below the estimate presented in the last section. The region-specific semi-elasticities are also similar: 1.85 for Canada, -0.06 for the US, 0.60 for the EU, 0.15 for the OECD, and 1.99 for emerging countries.

The second robustness check involves unweighted regressions, which means that studies presenting many estimates wield more influence in the meta-analysis. Table 2.7 shows that the posterior inclusion probabilities differ from the baseline specification for some variables. Concerning data characteristics, the age of data seems to be important: the reported semi-elasticity decreases each year by about 0.025. Studies that do not have direct data on within-country trade flows report larger estimates of the border effect. Adding one to zero trade flows typically yields lower semi-elasticities (by about 0.7). Moreover, the impact factor of the journal and the number of citations of the study seem to be important: better journals tend to re-

Table 2.6: Robustness check—alternative priors for BMA

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of Home	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value
Data characteristics						
Midyear of data	0.003	0.003	0.466	-0.001	0.012	0.926
Panel data	0.004	0.062	0.102			
Disaggregated	0.745	0.143	1.000	0.545	0.306	0.075
Obs. per year	0.000	0.008	0.060			
No. of years	0.113	0.082	0.738	0.100	0.098	0.310
Countries examined						
Canada	0.724	0.126	1.000	0.823	0.317	0.010
US	-1.183	0.133	1.000	-1.131	0.227	0.000
EU	-0.518	0.161	0.995	-0.548	0.383	0.152
OECD	-0.975	0.176	1.000	-0.902	0.343	0.009
Emerging	0.868	0.268	0.990	0.602	0.322	0.062
Design of the analysis						
No internal trade	0.184	0.209	0.508	0.361	0.389	0.354
Inconsistent dist.	0.754	0.145	1.000	0.521	0.304	0.087
Actual distance	-0.907	0.155	1.000	-0.716	0.331	0.030
Total trade	-0.001	0.062	0.041			
Asymmetry	0.518	0.121	0.999	0.492	0.246	0.045
Instruments	-0.008	0.054	0.055			
Treatment of multilateral resistance						
Remoteness	-0.016	0.066	0.090			
Country fixed eff.	0.362	0.334	0.601	0.214	0.272	0.431
Ratio estimation	0.628	0.491	0.721	0.738	0.506	0.145
Anderson est.	0.389	0.376	0.579	0.162	0.308	0.599
No control for MR	0.961	0.314	1.000	0.641	0.297	0.031
Treatment of zero trade flows						
Zero plus one	0.004	0.033	0.050			
Tobit	-0.640	0.155	0.998	-0.600	0.321	0.062
PPML	-0.726	0.155	1.000	-0.860	0.529	0.104
Zeros omitted	-0.007	0.035	0.074			
Control variables						
Adjancency control	0.125	0.156	0.453	0.341	0.245	0.163
Language control	-0.001	0.022	0.046			
FTA control	-0.253	0.167	0.778	-0.466	0.321	0.147
Publication characteristics						
Published	0.346	0.103	0.986	0.276	0.272	0.311
Impact	0.021	0.045	0.230			
Citations	0.003	0.014	0.077			
Publication year	0.074	0.011	1.000	0.055	0.032	0.083
Constant	0.081	NA	1.000	1.267	1.135	0.264
Studies	61			61		
Observations	1,271			1,271		

Notes: Home = the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows. PIP = posterior inclusion probability. SD = standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.3. Standard errors in the frequentist check are clustered at both the study and data set level (the implementation of two-way clustering follows Cameron *et al.* 2011). In this specification we use the beta-binomial prior advocated by Ley & Steel (2009) (prior model probabilities are the same for all possible model sizes) and set the Zellner's g prior following Fernandez *et al.* (2001). More details on the BMA estimation are available in Table 2.9 and Figure 2.7. A detailed description of all variables is available in Table 2.3.

Table 2.7: Robustness check—unweighted regressions

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of Home	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value
Data characteristics						
Midyear of data	-0.025	0.003	1.000	-0.027	0.006	0.000
Panel data	0.215	0.165	0.695	0.283	0.155	0.069
Disaggregated	0.619	0.120	1.000	0.537	0.235	0.022
Obs. per year	0.060	0.054	0.617	0.105	0.127	0.407
No. of years	0.022	0.050	0.195			
Countries examined						
Canada	0.996	0.137	1.000	0.940	0.293	0.001
US	-1.655	0.181	1.000	-1.730	0.285	0.000
EU	-1.317	0.114	1.000	-1.313	0.258	0.000
OECD	-1.069	0.159	1.000	-1.062	0.263	0.000
Emerging	0.870	0.164	1.000	0.810	0.233	0.001
Design of the analysis						
No internal trade	1.239	0.164	1.000	1.128	0.283	0.000
Inconsistent dist	0.016	0.071	0.074			
Actual distance	-0.655	0.215	0.970	-0.722	0.301	0.016
Total trade	0.005	0.056	0.030			
Asymmetry	0.001	0.023	0.028			
Instruments	-0.007	0.055	0.038			
Treatment of multilateral resistance						
Remoteness	-0.001	0.028	0.026			
Country fixed eff.	-0.002	0.044	0.040			
Ratio estimation	0.035	0.111	0.125			
Anderson est.	0.001	0.039	0.026			
No control for MR	0.489	0.131	0.990	0.470	0.177	0.008
Treatment of zero trade flows						
Zero plus one	-0.686	0.181	0.986	-0.571	0.308	0.064
Tobit	-0.131	0.221	0.309	-0.436	0.252	0.084
PPML	-0.969	0.174	1.000	-1.024	0.388	0.008
Zeros omitted	-0.001	0.025	0.028			
Control variables						
Adjancency control	0.093	0.147	0.336	0.294	0.221	0.184
Language control	-0.001	0.021	0.029			
FTA control	-0.015	0.062	0.083			
Publication characteristics						
Published	-0.001	0.032	0.031			
Impact	-0.186	0.055	0.979	-0.188	0.125	0.131
Citations	0.182	0.047	0.992	0.173	0.106	0.103
Publication year	0.097	0.015	1.000	0.089	0.039	0.023
Constant	2.750	NA	1.000	2.678	0.974	0.006
Studies	61			61		
Observations	1,271			1,271		

Notes: Home = the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows. PIP = posterior inclusion probability. SD = standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.3. Standard errors in the frequentist check are clustered at both the study and data set level (the implementation of two-way clustering follows Cameron *et al.* 2011). In this specification we do not weight regressions by the inverse of the number of estimates reported per study. More details on the BMA estimation are available in Table 2.10 and Figure 2.8. A detailed description of all variables is available in Table 2.3.

port smaller estimates, while broadly cited studies usually report larger estimates. Nevertheless, the best practice estimates of the border effect for the entire world and for individual regions are again very close to our baseline results, as shown in the right-hand part of Table 2.5. The overall mean semi-elasticity is 0.93, implying a partial equilibrium border effect of 2.5.

2.6 Criticisms of Meta-Analysis

In this section we list potential problems of conducting meta-analysis in economics and discuss how we address them. We identify 13 claims about meta-analysis that may cast doubt on the method:

1. *Studies of low quality should be excluded.* Our data set includes estimates from studies published in top journals, but also from studies not published in good outlets. As an alternative to meta-analysis, Slavin (1995) proposes "best evidence synthesis," which would only take into account the good studies. The obvious problem is where to draw a line between the good and the bad ones. We prefer to include as many papers as possible and give weight to different aspects of study design according to what we believe is the consensus on best practice methodology. In this way we can explore the influence of different methods on the estimated border effects. We also control for the impact factor of the publication outlet and for the number of citations each study gets.

2. *The analysis omits some studies.* We try to include as many studies as possible, but may still miss some. To allow other researchers to replicate our analysis, we use the query described in Section 2.2 to search for studies estimating the border effect. We believe it is not a problem to miss some studies, as long as their results do not differ systematically from the results of the included studies. With 1,271 estimates taken from 61 studies, our paper ranks among the largest meta-analyses conducted in economics (according to the survey by Doucouliagos & Stanley 2013).

3. *Studies reporting many estimates dominate the meta-analysis.* When each estimate gets the same weight, the unbalanced nature of data in meta-analysis means that studies with many estimates drive the results. One remedy involves the mixed-effects multilevel model, which gives each study approximately the same weight if within-study correlation of estimates is large (Havranek & Irsova 2011). The problem is that the method introduces study-level random effects, which may be correlated with explanatory variables. With so many explanatory variables defined at the study level, we prefer to simply weight regressions by the inverse of the number of estimates reported per study.

4. *Authors' preferred estimates should get more weight.* Studies examining the border effect usually present many estimates, and often prefer a subset of these estimates (many results are shown as robustness checks). Some authors make it clear what their preference is, but for many studies it is impossible to select the preferred estimates. We control for data and methodology instead, which is easier to code and should capture most preferences of the authors; for example, the control for multilateral resistance.

5. *Individual estimates are not independent, because authors use similar data.* Meta-analysis was originally designed for synthesizing medical research, where individual clinical trials can be considered approximately independent. In contrast, the regression results reported in economics are not independent, but neither are observations in most economics data sets. To account for the dependence among observations we cluster standard errors at the level of individual studies and data sets.

6. *Weighting by precision is inappropriate in economics because some methods underestimate standard errors.* Meta-analysts often use precision weights to remove heteroskedasticity in the regression estimating publication bias. We find no evidence of publication bias, so we can exclude the standard error from the equation and do not have to weight estimates by precision to yield efficiency.

Section 2.3 also illustrates that weighting by precision has little effects on the estimated border effect.

7. *Standard errors are not exogenous to the estimated coefficients.* When the choice of method systematically affects both the magnitude of the estimated border effect and its standard error, the explanatory variable in (2.2) will be correlated with the error term. Our solution is to use the number of observations as an instrument for the standard error: studies with more observations yield more precise estimates, but the number of observations is little correlated with the choice of methodology.

8. *The analysis omits some factors that may cause heterogeneity in the reported estimates.* We collect 32 aspects of data, methodology, and studies that may affect the estimated border effects. More specifics of study design could be included: for example, the exact method how internal distance is computed (we only include a dummy variable which equals one if the method differs from the computation of external distance); but we have to draw a line somewhere for the data collection to be feasible. Still we collect more variables that most meta-analyses in economics. Nelson & Kennedy (2009) review 140 meta-analyses and report that a median analysis uses 12 explanatory variables; the largest meta-analysis has 41 variables.

9. *There are too many potential explanatory variables and it is not clear which should be included.* With so many aspects of study design one cannot find a theory that would motivate the inclusion of all of them. For example, we would like to give more weight to large studies published in good journals, but it is not obvious why they should report systematically different results. We prefer to collect as many variables as possible and use Bayesian model averaging to resolve the resulting model uncertainty. The variables picked by BMA contain the ones that we feel should be included, such as the control for multilateral resistance and the measurement of internal distance.

10. *Meta-analysis compares apples with oranges.* Meta-analysis in economics examines heterogeneous estimates. Different estimates are produced using different methods, and we try to control for the differences in the design of primary studies. We also provide separate results for the regions examined in the literature. To increase the comparability of the estimates in our data set, we choose to only include the results concerning the effect of *international* borders on trade and omit the large literature on intranational border barriers.

11. *Meta-analysis may disagree with large primary studies.* The major reason for conducting meta-analyses in medical science is to increase statistical power by combining the small but costly clinical trials. Because individual clinical trials use similar methods, a comparison of meta-analysis with a later, large clinical trial provides a viable test of the reliability of meta-analysis. In economics the methods differ, and meta-analysis can be thought of as a weighted average of many different approaches. It would be difficult to construct a primary study that would reflect all recent advances in the methodology of the gravity equation, all possible aspects of our definition of best practice.

12. *Mistakes in data coding are inevitable.* The collection of data for meta-analysis involves months of reading and coding the data. We do not use research assistants for this work, because it is too tempting to jump directly to regression tables and code the data without reading much of the primary studies. We cannot exclude errors, but we do our best to minimize their number by collecting the data independently and then comparing and correcting the data sets.

13. *Publication bias invalidates meta-analysis.* When researchers prefer to report estimates showing a particular sign or statistical significance, the mean reported estimate will get biased. We test for publication bias in Section 2.3 and find little evidence of preferential selection. When we correct for any potential publication bias, we obtain a border effect close to the simple mean and median estimate. In general the file drawer problem matters for any type of literature synthesis, but meta-analysis can correct for the bias.

2.7 Concluding Remarks

We conduct a meta-analysis of the effect of international borders on trade. Using 1,271 estimates from 61 studies and controlling for differences in study quality, we show that the available empirical evidence suggests a mean reduction of 28% in international trade due to borders. The innovations introduced in the last decade to estimating the gravity equation alleviate the border puzzle worldwide and solve it for most OECD countries. Nevertheless, even after controlling for the advances in methodology we obtain large border effects for transition and developing countries.

To our knowledge, the only other quantitative survey on this topic is presented by Head & Mayer (2014, pp. 160–165), who compute the mean and median reported estimates of several important coefficients in the gravity equation, including the home coefficient. They collect 279 estimates from 21 studies and compute a mean and median home coefficient close to 2; in contrast, we find a mean and median close to 3. They focus primarily on studies published in top journals, while we gather more studies and control for study quality. Furthermore, Head & Mayer (2014) also collect estimates of the regression coefficient for the "same nation dummy," which serves as a control variable in many applications focusing on other issues than the border effect: for example, the trade effect of currency unions.

The same nation dummy usually has little variation and in most cases captures trade flows between large countries and their territories, such as between France and its overseas departments. The estimated coefficient for the dummy is often statistically insignificant and close to zero (see, for example, the results presented in Rose 2004), which is the primary reason why Head & Mayer (2014) obtain a smaller mean border effect than we do. They also include estimates of intranational home bias (for example, Wolf 2000), which we prefer to exclude and focus on the effect of international borders. In consequence only 10 primary studies overlap in the two meta-analyses.

Head & Mayer (2014) do not explicitly explore the heterogeneity in the estimates, but compute separate summary statistics for studies that control for multilateral resistance. For these studies they report a mean home coefficient of 1.9 and a median of 1.6. That is, Head & Mayer (2014) also find that the disregard of multilateral resistance exaggerates the estimated home coefficient, but their meta-analysis indicates that the bias is less than 0.4. Our results suggest that this aspect of methodology is more important: the omission of multilateral resistance terms biases the home coefficient by about 0.8, or about a quarter of the effect reported by McCallum (1995). In addition, we stress the importance of data aggregation, heterogeneity across regions, measurement of internal and external distance, and the treatment of zero trade flows.

References

Anderson, J. E. & E. van Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* **93(1)**: pp. 170–192.

Anderson, J. E. & E. van Wincoop (2004): "Trade Costs." *Journal of Economic Literature* **42(3)**: pp. 691–751.

Anderson, J. E. & Y. V. Yotov (2010): "The Changing Incidence of Geography." *American Economic Review* **100(5)**: pp. 2157–86.

Anderson, M. & S. Smith (1999): "Canadian Provinces in World Trade: Engagement and Detachment." *Canadian Journal of Economics* **32(1)**: pp. 22–38.

Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* **6(4)**: pp. 453–470.

Baldwin, R. & D. Taglioni (2007): "Trade Effects of the Euro: a Comparison of Estimators." *Journal of Economic Integration* **22**: pp. 780–818.

Balistreri, E. J. & R. H. Hillberry (2007): "Structural estimation and the border puzzle." *Journal of International Economics* **72(2)**: pp. 451–463.

Braconier, H. & M. Pisu (2013): "Road Connectivity and the Border Effect: Evidence from Europe." *OECD Economics Department Working Papers 1073*, OECD.

Cameron, A. C., J. B. Gelbach, & D. L. Miller (2011): "Robust Inference With

Multiway Clustering." *Journal of Business & Economic Statistics* **29(2)**: pp. 238–249.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-Analysis." *American Economic Review* **85(2)**: pp. 238–43.

Chetty, R., A. Guren, D. Manoli, & A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* **101(3)**: pp. 471–75.

Dekle, R., J. Eaton, & S. Kortum (2007): "Unbalanced Trade." *American Economic Review* **97(2)**: pp. 351–355.

DeLong, J. B. & K. Lang (1992): "Are All Economic Hypotheses False?" *Journal of Political Economy* **100(6)**: pp. 1257–72.

Disdier, A.-C. & K. Head (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics* **90(1)**: pp. 37–48.

Doucouliagos, H. & T. D. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Egger, M., G. D. Smith, M. Scheider, & C. Minder (1997): "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* **316**: pp. 629–634.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Feenstra, R. C. (2002): "Border Effects and the Gravity Equation: Consistent Methods for Estimation." *Scottish Journal of Political Economy* **49(5)**: pp. 491–506.

Feldkircher, M. & S. Zeugner (2009): "Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging." *IMF Working Papers 09/202*, International Monetary Fund.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

Görg, H. & E. Strobl (2001): "Multinational Companies and Productivity Spillovers: A Meta-analysis." *The Economic Journal* **111(475)**: pp. F723–39.

Havranek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85(2)**: pp. 234–244.

Head, K. & T. Mayer (2000): "Non-Europe: The magnitude and causes of market fragmentation in the EU." *Review of World Economics* **136(2)**: pp. 284–314.

Head, K. & T. Mayer (2010): "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade." In P. A. G. van Bergeijk & S. Brakman (editors), "The Gravity Model in International Trade: Advances and Applications," pp. 165–192. Cambridge University Press.

Head, K. & T. Mayer (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook." In "Handbook of International Economics," volume 4, pp. 131–195. Elsevier.

Hillberry, R. H. (2002): "Aggregation bias, compositional change, and the border effect." *Canadian Journal of Economics* **35(3)**: pp. 517–530.

Kee, H. L., A. Nicita, & M. Olarreaga (2009): "Estimating Trade Restrictiveness Indices." *Economic Journal* **119(534)**: pp. 172–199.

Ley, E. & M. F. Steel (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions." *Journal of Applied Econometrics* **24(4)**: pp. 651–674.

McCallum, J. (1995): "National Borders Matter: Canada-U.S. Regional Trade Patterns." *American Economic Review* **85(3)**: pp. 615–23.

Nelson, J. & P. Kennedy (2009): "The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment." *Environmental & Resource Economics* **42(3)**: pp. 345–377.

Nitsch, V. (2000): "National borders and international trade: evidence from the

European Union." *Canadian Journal of Economics* **33(4)**: pp. 1091–1105.

Obstfeld, M. & K. Rogoff (2001): "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" In "NBER Macroeconomics Annual 2000, Volume 15," NBER Chapters, pp. 339–412. National Bureau of Economic Research, Inc.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* **92**: pp. 179–191.

Rose, A. K. (2000): "One money, one market: the effect of common currencies on trade." *Economic Policy* **15(30)**: pp. 7–46.

Rose, A. K. (2004): "Do We Really Know That the WTO Increases Trade?" *American Economic Review* **94(1)**: pp. 98–114.

Rosenthal, R. (1979): "The 'File Drawer Problem' and Tolerance for Null Results." *Psychological Bulletin* **86**: pp. 638–41.

Rusnak, M., T. Havranek, & R. Horvath (2013): "How to Solve the Price Puzzle? A Meta-Analysis." *Journal of Money, Credit and Banking* **45(1)**: pp. 37–70.

Siegfried, J. J. (2012): "Minutes of the Meeting of the Executive Committee: Chicago, IL, January 5, 2012." *American Economic Review* **102(3)**: pp. 645–52.

Silva, J. M. C. S. & S. Tenreyro (2006): "The Log of Gravity." *The Review of Economics and Statistics* **88(4)**: pp. 641–658.

da Silva, O. M., F. M. de Almeida, & B. M. de Oliviera (2007): "Comércio internacional "x" intranacional no Brasil: medindo o efeito-fronteira." *Nova Economia* **17(3)**: pp. 427–439.

Slavin, R. E. (1995): "Best evidence synthesis: an intelligent alternative to meta-analysis." *Journal of Clinical Epidemiology* **48(1)**: pp. 9–18.

Stanley, T. & H. Doucouliagos (2010): "Picture This: A Simple Graph That Reveals Much Ado About Research." *Journal of Economic Surveys* **24(1)**: pp.

170–191.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-Analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

Wei, S.-J. (1996): "Intra-National versus International Trade: How Stubborn are Nations in Global Integration?" *NBER Working Papers 5531*, National Bureau of Economic Research, Inc.

Wolf, H. C. (2000): "Intranational Home Bias In Trade." *The Review of Economics and Statistics* **82(4)**: pp. 555–563.

2.A Diagnostics of BMA

Table 2.8: Summary of BMA estimation, baseline specification

Mean no. regressors 18.5374	*Draws* $2 \cdot 10^6$	*Burn-ins* $1 \cdot 10^6$	*Time* 6.914583 minutes
No. models visited 311, 863	*Modelspace* $4.3 \cdot 10^9$	*Visited* 0.0073%	*Topmodels* 98%
Corr PMP 0.9994	*No. Obs.* 1, 271	*Model Prior* uniform	*g-Prior* UIP
Shrinkage-Stats Av= 0.9992			

Notes: In this specification we employ the priors suggested by Eicher *et al.* (2011) based on predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of data).

Figure 2.6: Model size and convergence, baseline specification

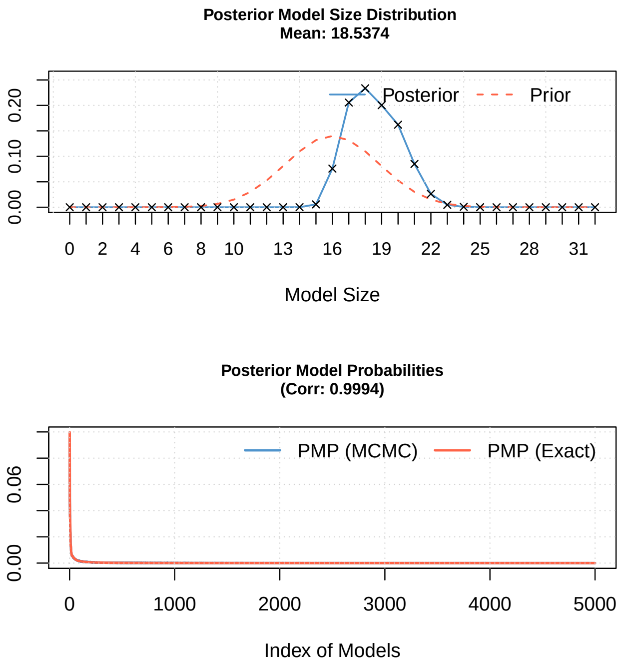


Table 2.9: Summary of BMA estimation, alternative priors

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
19.6891	$2 \cdot 10^6$	$1 \cdot 10^6$	7.2395 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
$394,789$	$4.3 \cdot 10^9$	0.0092%	96%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9993	$1,271$	random	BRIC
Shrinkage-Stats			
Av= 0.9992			

Notes: The "random" model prior refers to the beta-binomial prior advocated by Ley & Steel (2009): prior model probabilities are the same for all possible model sizes. In this specification we set the Zellner's g prior following Fernandez *et al.* (2001).

Figure 2.7: Model size and convergence, alternative priors

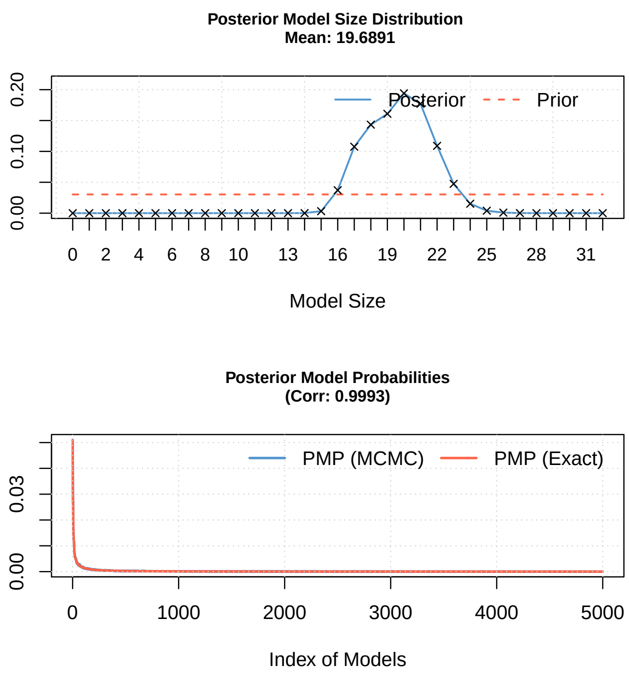
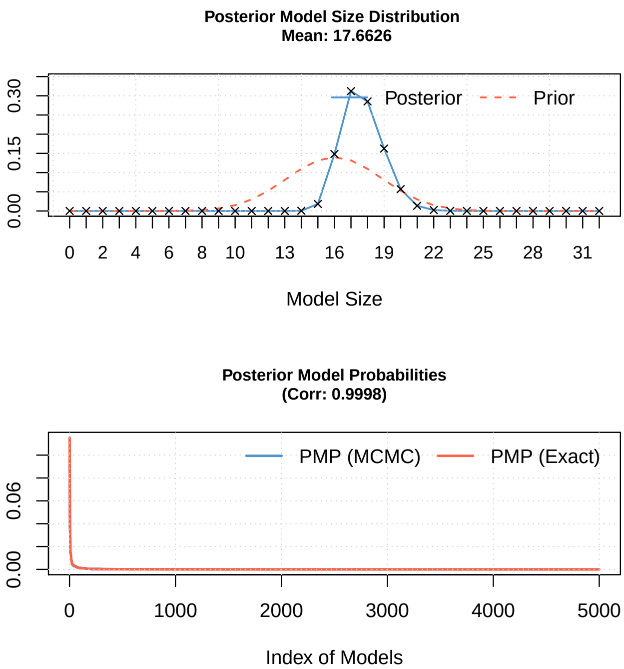


Table 2.10: Summary of BMA estimation, unweighted regressions

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
17.6626	$2 \cdot 10^6$	$1 \cdot 10^6$	7.121633 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
350,260	$4.3 \cdot 10^9$	0.0082%	98%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9998	1,271	uniform	UIP
Shrinkage-Stats			
Av= 0.9992			

Notes: In this specification we employ the priors suggested by Eicher *et al.* (2011) based on predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of data).

Figure 2.8: Model size and convergence, unweighted regressions



Chapter 3

Meta-Analysis of Intra-Industry FDI Spillovers: Updated Evidence

Abstract

We conduct a meta-analysis of literature on intra-industry productivity spillovers from foreign direct investment using 67 published and unpublished studies. Combined significance of individual t statistics is inconclusive but it is apparent that papers published in leading journals tend to report rather insignificant results. Meta-regression analysis confirms that cross-sectional and industry-level studies are likely to find relatively strong spillover effects. Moreover the choice of proxy for foreign presence is important. The pattern, however, seems to weaken over time. Evidence for publication selection bias was detected employing the funnel asymmetry test, and the spillover effect corrected for publication bias is not significantly different from zero.

Keywords: Meta-analysis, productivity spillovers, foreign direct investment, multinational companies

JEL Codes: C83, D62, F21

This paper is a joint work with Tomas Havranek. We thank Vladimir Benacek, Tomas Cahlik, Pekka Ilmakunnas, Ladislav Kristoufek, participants of the AAEM seminar at Charles University in Prague, and two anonymous referees of this journal for valuable comments. We acknowledge support from the IES Research Institutional Framework 2005–2010 (MSMT 0021620841). The paper has been published in the Czech Journal of Economics and Finance.

3.1 Introduction

Governments all over the world pay fortunes, either in cash or as tax holidays, to attract inward foreign direct investment (FDI) under their jurisdiction. There are many reasons why governments attempt to lure multinational companies (MNCs) but the principal one resides in their expectations of positive productivity externalities spilling over from MNCs to domestic firms (see Blomström & Kokko 2003). A substantial body of empirical literature on productivity spillovers has been published since the 1970s and many narrative literature reviews were conducted (see, *inter alia*, Pack & Saggi 1997). The first quantitative survey, commonly called a meta-analysis, was conducted by Görg & Strobl (2001), followed by Meyer & Sinani (2005), and Wooster & Diebel (2006). For a discussion of the advantages and disadvantages of narrative and quantitative methods of literature reviews, see Stanley (2001).

Meta-analysis is a rather new method in economics; it has been employed only since the 1980s, and the meta-regression approach, which is the main focus of this paper, was developed by Stanley & Jarrell (1989). The recent economic research by means of meta-analysis covers for instance Gallet (2007) trying to uncover the extent to which study characteristics influence the estimates of tuition and income elasticities, Li *et al.* (2007) investigating systematic variation across environmental Kuznets curve studies, Fidrmuc & Korhonen (2006) who present a study on business cycle correlation between the Euro area and the Central-East European Economies, or Havránek (2010) investigating the effect of currency unions on international trade.

A meta-analyst rigorously combines the outcomes of several works that study the same phenomena. A meta-regression analyst, in the concrete, collects a number of statistics from the targeted literature—e.g., correlation coefficients or t statistics of estimates of the effect in question—and regresses it on several proxies of the study design. If any of meta-explanatory variables is found to be significant, it is taken as an evidence that studies' results are systematically dependent on their design (for a good introduction to the meta-regression technique, see Stanley 2001). Concerning the meta-analyses of the spillover literature, Görg & Strobl (2001) apply plain ordinary least squares (OLS) meta-regression, Meyer & Sinani (2005) employ panel data methods, and Wooster & Diebel (2006) perform logistic meta-regression. We combine all the three methods and include also robust estimations to check sensitivity of the results. The sample of literature used in this meta-analysis is also much broader than in the previous analyses, containing 67 original empirical works. Most importantly, compared to previous studies, we apply the modern methodology of correcting for publication bias (Doucouliagos & Stanley 2009; Doucouliagos & Laroche 2009) on the spillover literature.

The present paper is structured as follows: Section 3.2 lists channels of transfers of intra-industry (or horizontal) spillovers from MNCs to domestic firms, discusses spillover determinants, and describes the standard design of empirical works on horizontal spillovers. Section 3.3 discusses in detail the literature selection procedure which was employed and describes properties of the resulting data set. Section 3.4 investigates the combined significance of the collected t statistics. In Section 3.5 the meta-regression analysis is performed. Section 3.6 tests for the presence of publication bias in the spillover literature. Section 3.7 concludes.

3.2 Horizontal Spillovers from FDI

The history of intra-industry productivity spillover literature[1] dates from 1960, covering works of MacDougall (1960), Corden (1974), or Caves (1974), who analyzed the welfare effects of FDI, its impact on optimal tariff policy, industrial level, and international trade openness. A deeper specification is provided in Blomström & Kokko (1996), embodied in the three main channels of technology transfer:

Competition effect The entry of foreign enterprises contributes to the development on industrial, technological, and managerial level and export dynamics through the creation of competitive environment. Nevertheless, multinational companies may

[1] Testing for vertical spillovers following the seminal work of Javorcik (2004) has become popular in the recent literature. This article is, however, focused on horizontal spillovers. Other FDI externalities than productivity spillovers have been discussed as well, specifically the market access spillovers (see, e.g., Blomström & Kokko 2003) or financing spillovers (Geršl & Hlaváček 2007; Geršl 2008), but there are only a few empirical studies estimating those.

evoke crowding-out effects, generating harmful externalities to the domestic firms. MNCs can acquire significant market shares, reducing the opportunities of domestic firms to exploit returns to scale (Aitken & Harrison 1999) or drain scarce resources. Such detrimental effects of FDI are highlighted by several researchers (for instance, Haddad & Harrison 1993, who, in fact, present evidence of *negative* horizontal spillovers).

Demonstration effect Realization of the demonstration effect stems from the differences in technology level between foreign investors and host-country firms. MNCs enter the host-country market and establish affiliates which possess superior technology compared to the local companies. Locals observe and imitate these affiliates in the same industry, thus becoming more productive. In some cases only a direct contact with new technologies can overcome conservative attitudes toward the implementation of up-to-date technologies (Blomström & Kokko 1996).

Labor turnover Host country's citizens employed by the foreign investor might benefit from the contact with advanced technologies and production methods. Based on the transfer of human capital, knowledge, and skills toward the host country labor force, this labor exchange phenomena can enhance competitiveness of domestic firms. MNCs train local labor force since it is still cheaper than to import skilled labor from their home country; even though, in most cases, they cannot prevent the labor turnover (see Görg & Greenaway 2004).

Researchers have been recently turning their attention toward the question of real spillover heterogeneity. It was shown that the existence, polarity, and magnitude of FDI spillovers depends on various factors related especially to MNCs, domestic firms, and regional characteristics (for a comprehensive survey of spillover determinants, see Crespo & Fontoura 2007).

The most frequently pronounced determinant is the absorptive capacity of domestic firms—the ability to adopt new technologies from MNCs. Technological gap is often employed to approximate this determinant: the importance of FDI spillovers is maximized if the technological gap is moderate; not too high but also not too low (Kokko 1994). As another proxy, R&D expenditures are also used (Griffith *et al.* 2003). At the macro level, many authors associate the absorptive capacity with the host country's development (see Xu 2000, or Lipsey & Sjöholm 2004 in context of labor channel: developing countries may benefit less from labor turnover since it is difficult for the domestic firms to offer wages competitive with the MNCs), human capital (more advanced technology is connected with higher proportion of skilled labor, Blomström *et al.* 1994, Kokko & Blomström 1995) or developed financial system (which reduces investment risk of domestic firms willing to adopt new technologies, Hermes & Lensink 2003).

The theory behind the regional effects, another major determinant, suggests that FDI spillovers decrease with an increasing geographical distance between domestic firms and MNCs (Audretsch 1998) since the three channels described above are limited in space (Girma 2003; Torlak 2004). Additionally, there are various determinants related to the domestic firms characteristics. With relation to the firms export capacity, FDI spillovers may be higher for non-exporting domestic firms since the exporting ones already face sufficient competition pressures (Blomström & Sjöholm 1999), domestic market is less relevant for them and they are more experienced with foreign competition (Barrios & Strobl 2002, Schoors & Tol 2002); therefore the entry of MNCs would not cause high competition effects for such companies. Another factor is the firm size; small firms not being able to profit from returns to scale may be less prepared to compete with MNCs. Concerning different types of firms the private or state ownership, for example, may influence the firms' absorptive capacity (see Sinani & Meyer 2004).

The other important factors include trade policy environment (with inward-oriented policy, for example, MNCs are likely to use technologies unknown to domestic firms, Kokko *et al.* 2001). Lee & Mansfield (1996) suggest that the level of protection of intellectual property rights increases the probability that MNCs would use more advanced technology, augmenting the magnitude of spillovers. Fosfuri *et al.*

(2001) find evidence for the importance of the type of training received by workers at MNCs and the existence of restrictions on labor mobility. According to Wang & Blomstrom (1992), MNCs facing strong competition would use more advanced technology; on the other hand, Fosfuri *et al.* (2001) argues that this would lead MNCs to protect their know-how more carefully.

Many of the firm-level determinants have been tested in the spillover literature and found significant for the magnitude and polarity of productivity spillovers. While in this article we search for methodological spillover determinants (i.e. how the methodology chosen by the researcher can systematically influence the reported results), it is important to keep in mind that there might exist some real heterogeneity in the spillover literature.

Since it is not possible the measure the three channels of technology transfer directly, empirical works on horizontal productivity spillovers are usually performed in the following way: researchers collect data on firms' productivity or output (either on firm or industry level) and regress it on a measure of foreign presence in the firms' industries, controlling also for additional variables (capital/output, labor/output ratios, etc.). If the estimate of the parameter for foreign presence is found to be positive and significant, the authors conclude that there is some statistical evidence for the existence of intra-industry spillovers.

3.3 The Sample of Literature

In the present paper, 97 results from 67 different studies are used, which is a significant increase compared to Görg & Strobl (2001), who used a sample of 21 studies, or Meyer & Sinani (2005) and Wooster & Diebel (2006), who had at their disposal 41 and 32 studies, respectively. We tried to include all relevant papers listed in the previous meta-analyses; additional search was performed in the EconLit, RePEc, and Google Scholar databases using combinations of the keywords "foreign direct investment", "productivity spillovers", and "technology transfer".

We follow the approach of Görg & Strobl (2001) in the selection process, i.e., only those studies are included which do not diverge significantly from the standard methodology of productivity-spillovers empirical work as it is described in Section 3.2, and only English-written papers are considered. No pre-selection for quality was employed. In the first place, we do not use results for inter-industry (or vertical), innovation, market access, and financing spillovers. These categories are qualitatively relative, but the tested specifications are, in our opinion, too dissimilar to be pooled together in the framework of a meta-analysis, and it would be much more appropriate to analyze such streams of literature separately. The more distant models are used the more heterogeneous the sample becomes and the less reliable are the results drawn from it. Random-effects meta-analysis may provide a remedy for heterogeneity (see, *inter alia*, Hedges 1992), but better approach may be to try to avoid the problem as much as possible.

Excluding inter-industry, innovation, market access, and financing spillovers, there is still a substantial body of empirical literature dealing with horizontal productivity spillovers. Many papers present multiple models, and thus multiple results. As a rule, we tried to choose the one that was considered the best by the researchers themselves. If the preferred model was not suitable for the analysis, i.e., it diverged too much from the standard methodology, the model with the highest R-squared (or adjusted R-squared, depending on which one was published) was selected. There are also works that examine different countries with the same methodology, or one country with different specifications which are, nevertheless, consistent with the mainstream approach (such estimates are called "conceptually independent" in the meta-analysis literature). For example, Konings (2000) studies spillovers in Bulgaria, Poland, and Romania separately, thus 3 observations were included from his paper. Liu (2008) first presents a purely firm-level model but subsequently adds industry dummies, thus we obtain two observations from this paper, etc. On the other hand, Sadik & Bolbol (2001) apply not industry- or firm-, but country-level aggregation, and Zhu & Tan (2000) uses city-level data set, therefore we do not include these papers—although Wooster & Diebel (2006) use them. Rattsø & Stokke (2003) employ two proxies for foreign presence at the same time, the share of trade on GDP and FDI on overall investment, none of them belonging to the standard measures in the spillover literature—thus this paper is also excluded from the meta-analysis. It is difficult to model idiosyncratic research choices in the meta-regression framework.

We realize that the selection process is the most vulnerable part of the present work, but the final sample is broad and represents works of researchers from dozens of countries and evidence from many economies around the world. Both journal articles and working papers were used. The list of studies and some of their characteristics can be found in Table 3.7 in Section 3.A.

The first aspect of the study design that we include in the meta-analysis is the status of the country for which the data are used. From the whole sample of 97 observations, 41 models are using data for developing countries, 34 models' data are for transition countries, and 22 for advanced economies. Countries are distributed in groups according to the European Economic Association (transition countries list) and the World Bank (developing economies list) as of 2008. The second aspect is the (non)existence of time dimension in the data. Thirty-two models use cross-sectional data, the remaining 65 models rely on panel-data techniques. The third aspect is the definition of MNCs' presence. Thirty-two specifications define foreign presence in the industry as foreign firms' share on employment, 25 use assets, 21 output (or value added), and 19 share on sales. The fourth aspect is the level of aggregation. Forty models use purely firm-level data, whereas 35 include also industry dummies and 22 aggregate data on the level of industries. The fifth aspect is the definition of the response variable. Thirty-nine specifications use output growth, 54 models apply labor (or total factor[2]) productivity level or log-level and the rest employ other measures (for details of different measures, see Görg & Strobl 2001). Exact definitions of all variables and their summary statistics can be found in Table 3.8 in Section 3.A.

[2]To simplify, we abstract from the fact that there are different ways how to estimate total factor productivity, although it might also affect the extent of detected spillovers.

3.4 Combined Significance

Once we have collected a broad sample of empirical studies on intra-industry spillovers, the most natural question appears to be: can we somehow decide whether or not there is any general evidence for the existence of the spillover effect? The crucial result of every empirical work on productivity spillovers is the (non)significance, polarity, and magnitude of the estimate of the regression parameter which corresponds to the variable that is used as a proxy for foreign presence in the industry. Since every researcher can (and generally does) use different units, it is not appropriate to take the magnitude of estimates as the representative variable. The t statistic, on the other hand, is a dimension-less variable which is widely employed for the purposes of a meta-analysis (it is also used by all three existing meta-analyses of the spillover literature: Görg & Strobl 2001; Meyer & Sinani 2005; Wooster & Diebel 2006).

The first possible way how to evaluate combined significance is to employ the so-called "vote-counting method" (see, *inter alia*, Hunter & Schmidt 1990). Following this approach, one would count the median value of t statistics in the sample; let us denote it T_M. If the median value was significant, this could be taken as an evidence for the existence of the phenomenon in question, and *vice versa*. This method has been criticized, e.g., by Djankov & Murrell (2002). Instead of the vote-counting method, they examine the following statistics:

$$T = \frac{\sum_{k=1}^{K} t_k}{\sqrt{K}}, \tag{3.1}$$

where K denotes the number of models included in the meta-analysis (i.e., $K = 97$ in our case) and t_k is the t statistic taken from the k-th model. Provided that all studies are independent and have sufficiently large number of degrees of freedom, T is normally distributed and combined significance can be easily tested. Note that, from this point of view, the vote-counting method drastically under-values the "real" effect. Indeed, many meta-analysts (e.g., Hedges & Olkin 1985) consider it to be obsolete. Still, it is widely used especially in narrative literature reviews.

Table 3.1: Aggregated t statistics

Variable	**# of obser.**	**Without outliers**	**All studies** T	T_W	T_M	**Without outliers** T	T_W	T_M
all	97	87	15.5	10.1	0.4	4.41	3.99	0.3
developing	41	38	6.95	2.11	0.9	4.54	3.27	0.811
transition	34	31	9.12	8.13	−0.00423	−0.569	−0.645	−0.193
advanced	22	18	11.8	13.6	1.4	3.85	4.4	0.85
cs	32	28	21.3	16.4	2.77	10.4	9.28	2.41
panel	65	59	4.02	−1.05	0.000265	−1.8	−3.098	0.000185
empl	32	28	13.2	11.9	1.85	6.907	7.9	1.4
sales	19	16	3.102	−3.15	−0.323	−1.03	−1.98	−0.326
assets	25	23	6.81	3.55	0.0507	−0.417	−2.96	0.037
output	21	20	6.64	4.17	0.9	2.4	2.74	0.7
firm	40	35	6.031	0.491	0.312	0.214	−1.28	0.3
industry	22	21	10.8	9.57	2.41	9.22	8.17	2.4
secdum	35	31	10.8	10.4	0.000265	−0.42	0.477	$2.28 \cdot 10^{-6}$
growth	39	33	12.5	6.29	0.4	2.46	−0.358	0.324
prod	58	54	9.82	7.97	0.531	3.68	4.92	0.282
old	46	42	11.3	6.068	1	4.77	4.066	0.75
new	51	45	10.7	8.41	0.324	1.53	1.48	0.051
journal	32	28	17.4	16.2	1.42	6.19	5.64	0.811
wp	42	39	3.69	6.76	0.0258	2.086	4.73	0.000957
topjournal	23	20	6.35	1.43	0.99	−1.029	0.499	0.445

Notes: all stands for dataset including all available data; *developing* restricts overall dataset to developing countries only, *transition* to transitional countries, and *advanced* to advanced countries; *cs* is a subsample of studies using cross-sectional data and *panel* using panel data; *empl* is a subsample for studies measuring MNC presence in employment, *sales* in sales, *assets* in assets, and *output* in output; *firm* is a sample restricted to studies on firm-level, *industry* for industry-level, *secdum* if industry dummies are used; *growth* is a dummy variable equal to 1 if response variable is output growth and *prod* if labor productivity; *new* stands for studies published since 2003, *old* otherwise; *journal* restricts the sample to articles published in other journals than *topjournals*, *wp* stands for working papers, *topjournal* denotes articles published in the leading 60 economics journals.

Djankov & Murrell (2002) also propose another modification of (3.1):

$$T_W = \frac{\sum_{k=1}^{K} w_k t_k}{\sqrt{\sum_{k=1}^{K} w_k^2}}, \tag{3.2}$$

where w_k are weights assigned to the k-th model, T_W being normally distributed. Both (3.1) and (3.2) are used in meta-analyses of the spillover literature. Meyer & Sinani (2005) assign higher weights to the models that employ "sophisticated econometric methods", Wooster & Diebel (2006) simply use the inversion of the number of models taken from a particular paper (for example, if 3 models from the paper are taken, each has the weight 1/3). We define a combined weight which accounts for (i) the number of models taken from a particular paper as in Wooster & Diebel (2006), and (ii) the "quality" of the paper. Quality is proxied by the level of publication, i.e., working papers have the lowest weight ($w = 0.25$), articles published in lesser journals have moderate weight ($w = 0.5$), and articles published in the top 60 economics journals according to the list by Kalaitzidakis *et al.* (2003) have the full weight ($w = 1$). It would be possible to take more complicated weights, e.g., some distribution of impact factors, but then there would be a problem with weights for working papers. Nevertheless, even such simple weights have significant impact on the results, as can be seen from Table 3.1.

Table 3.1 shows the combined significance of the spillover effect in different groups of the sample. Both normally distributed statistics T (3.1) and T_W (3.2), and the median value T_M are reported. Values of t_k from our sample vary significantly, from the lowest point of -11.58 to the peak of 27.7. Because such excessive values have rather dramatic effect on the combined significance, we report also T, T_W, and T_M for a narrower sample without these outliers. More concretely, we employ the restriction $|t_k| \leq 8$, thus the narrower sample contains 87 observations. From these 6 measures of combined significance, we would prefer T_W without outliers. It is evident at first sight that the weighted value (T_W) is in most cases below the simple measure T, indicating that better-quality papers may report lower t statistics, or that discounting the weights for multiple models taken from one paper has a powerful effect. Nevertheless, for the pooled sample both T and T_W are highly significant, even with the exclusion of outliers. T_M, on the other hand, is not significant. To conclude, the spillover effect is, in general, not significant according to the vote-counting method, but it is significant applying the Djankov & Murrell (2002) methodology.

There are two groups in the sample for which the spillover effect is significant independently of the methodology in use or spillovers exclusion—these are studies using cross-sectional data and studies with industry-level aggregation. Specifications that measure MNCs' presence as a share of employment are together not significant only when the combined t statistics is measured by T_M without outliers. On the other hand, for firm-level specifications, panel data models, studies using sales as a measure of foreign presence, and papers published in the top 60 world economics journals, combined t statistics are positively significant only if they are measured simply as T and outliers are included; the remaining 5 measures are insignificant or even negatively significant. Based on this finding, one could argue that there might be a tendency in the most prestigious journals to publish rather skeptical empirical studies on productivity spillovers, or—perhaps more probably—that papers of high quality might be more probable to find no or even negative spillover effects. However, at first sight, it seems that the effect of quality on the results is not linear, since studies published in lesser journals are more likely to find positive spillovers than studies published only as working papers. But recent working papers can still be published in a journal, either top or not—thus the "mixed" results for working papers do not stand against our main argument. Based on several sensitivity checks, the present authors would argue that the trend among the most respected journals is obvious and that minor changes in the definitions of the top journals would not change the conclusion.

It is also interesting that for transition countries excluding outliers all three combined t statistics are insignificant and even negative. This can be surprising since transition countries are usually considered to be likely to benefit from FDI highly as, in their case, the technology gap between domestic firms and MNCs is moderate (see, e.g., Blomström & Kokko 2003). Furthermore, it seems that newer studies (those published after 2002, dividing the sample approximately to 2 halves) might be more probable to report insignificant results, although the effect of studies' age does not appear to be very strong.

3.5 Meta-Regression Analysis

We have already seen that various aspects of studies' design are likely to influence the result—which is the t statistic for the estimate of the coefficient that represents the measure of foreign presence in the industry. In this section, we would like to investigate this pattern more thoroughly, using a different approach known as the meta-regression analysis. As a benchmark case, we follow Görg & Strobl (2001) who run a plain OLS regression:

$$Y_k = \alpha + \sum_{l=1}^{L} \beta_l X_{kl} + \epsilon_k, \quad k = 1, 2, \ldots, K, \tag{3.3}$$

where the meta-response variable Y_k is the t statistic from the k-th specification and meta-explanatory variables X_{kl} reflect different aspects of studies' design according to the 5 main features from Section 3.3—i.e., those that can be chosen by the researchers *ex ante*.[3] For this reason, we do not include a dummy for the level of publication. Because in the absence of publication bias there should be a significant and positive relation between the number of degrees of freedom in the particular model and its reported (absolute) value of t statistic, the logarithm of degrees of freedom makes an additional meta-explanatory variable. Another aspect we would like to control for is the time period for which the study was conducted, thus we include the average year of study period as a meta-explanatory variable. The final model consist of 11 meta-explanatory variables for 97 observations, which gives us much more degrees of freedom than Görg & Strobl (2001) have (25 observations for 9 regressors).

Descriptions of all variables can be found in Table 3.8 in Section 3.A. First, we

[3] Baseline case: data are firm-level, panel, and for a developed country, response variable is specified in productivity level, log-level, or "other", foreign presence is measured in sales.

examine relationships between meta-explanatory variables. The table of correlation coefficients (Table 3.10) is included in Section 3.C, as well—the highest absolute value of all correlation coefficients, 0.63, does not seem to indicate multicollinearity. The condition number is high, but it is sufficient to exclude the average year of study period and it declines to 16. In the regression model, exclusion of this variable does not change the estimated signs neither the significances of estimates, thus we mostly work with the complete number of meta-explanatory variables. More discussion of multicollinearity can be found in Section 3.B.

All regressions were conducted in Stata 10 and the most important results employing different estimators are summarized in Table 3.2. Detailed results of the standard meta-regression using OLS are reported in Table 3.11 in Section 3.C. We found it necessary to exclude the most obscure observations—with $|t_k| > 8$. There are three main reasons for such selection. Firstly, observations with such a high absolute value of t statistic reach also the largest values of Cook's distance for specification 1 of Table 3.11 and their predicted residuals are high. Secondly, there is a large gap between the observation with the absolute value of t statistic equal to 5.9 and the next higher one 8.4. Thirdly, it is a similar cut-off level as was used by Görg & Strobl (2001). Nevertheless, we report both types of specifications (with and without outliers) in Table 3.11.

Performing standard tests of suitability of the model (referring to specification OLS of Table 3.2), the Ramsey RESET test does not reject the null hypothesis, thus the selected specification is not considered to be wrong. Results of multicollinearity analysis and analysis of non-linear relationships do not change when outliers are excluded. To deal with a possible presence of heteroscedasticity of disturbances, we use heteroscedasticity robust standard errors computed with the Huber-White sandwich estimator, see Huber (1967) and White (1980). To test for normality of disturbances, we employ the Shapiro-Wilk test, which rejects the null hypothesis. This is one of the reasons for which we decided to employ also other methods, not only plain OLS as Görg & Strobl (2001).

The most obvious choice is to use some of robust estimators, which can also help to assess whether the selected cut-off level for outliers in OLS was the right one. We decided for two alternative estimators, iteratively re-weighted least squares (IRLS) with Huber and Tukey bisquare weight functions tuned for 95% Gaussian efficiency (see Hamilton 2006, pp. 239–256) and median regression[4] from the family of quantile regressions. Results of the robust meta-regression can be found in Table 3.12 in Section 3.C. Concerning the selection of outliers in OLS, we can see that, e.g., IRLS predicts results that are very similar to that of OLS without outliers. Therefore we can conclude that the cut-off $|t_k| \leq 8$ does not seem to be improperly chosen.

Following Meyer & Sinani (2005), we also perform a pseudo-panel data meta-regression. The cross-sectional dimension is represented by different papers, the other dimension is the order of a model taken from a particular paper. Because we have 97 observations from 67 papers at our disposal, it would not be wise to use the fixed-effects model, as many observations would be dropped and the number of degrees of freedom would diminish significantly, thus it is not even possible to test for fixed effects reliably. Therefore, we will assume that the study-specific effect is normally distributed (nevertheless, this kind of extreme unbalancedness might have an effect on the random effects estimates as well). We will test the following unbalanced panel data model:

$$Y_{ij} = \alpha_i + \sum_{l=1}^{L} \beta_l X_{ijl} + \epsilon_{ij}, \quad i = 1, 2, \ldots, 67, \quad j = 1, 2, \ldots, 8. \tag{3.4}$$

Details of random-effects meta-regression are reported in Table 3.13 in Section 3.C. It is apparent that, excluding outliers, there is no substantial difference in the predictions of plain OLS and random-effects regression. Testing for random effects, the Breusch-Pagan Lagrange multiplier test does not reject the null hypothesis (it is significant only at the 15% level), thus it might suffice to perform plain OLS in this case. But there is one other advantage of the panel-data method: as Stanley (2001) remarks, if a meta-analyst takes a lot of observations from one paper, a single researcher (or even a single work) can dominate the whole meta-regression. This is

[4]The algorithm minimizes the sum of the absolute deviations about the median.

Table 3.2: Summary of conducted meta-regressions, all studies

Response variable: t statistic; dummy = 1 if positive (probit)	OLS	IRLS	Median reg.	RE	Probit
Logarithm of degrees of freedom	0.0969	0.137	0.100	0.0828	0.0666
	(0.69)	(1.06)	(0.78)	(0.60)	(0.79)
Average year of study period	−0.0119	−0.0216	−0.0239	−0.00560	−0.0195
	(−0.40)	(−0.62)	(−0.71)	(−0.18)	(−0.54)
Dummy = 1 if data are for developing country	−0.124	0.0353	−0.0411	−0.247	0.264
	(−0.23)	(0.05)	(−0.07)	(−0.47)	(0.53)
Dummy = 1 if data are for transition country	0.805	0.833	1.068	0.727	0.635
	(0.99)	(1.03)	(1.37)	(0.89)	(1.08)
Dummy = 1 if data are cross-section	2.023^{**}	1.876^{**}	2.363^{**}	1.993^{**}	1.123^{*}
	(3.16)	(2.91)	(3.70)	(3.10)	(2.46)
Dummy = 1 if response variable is output growth	$0.973^{\dagger}$	0.880	0.839	0.756	0.162
	(1.91)	(1.64)	(1.57)	(1.47)	(0.46)
Dummy = 1 if data are industry-level	1.851^{**}	1.884^{*}	0.770	1.787^{**}	1.602^{*}
	(2.85)	(2.37)	(1.03)	(2.74)	(2.27)
Dummy = 1 if industry dummies are used	0.237	0.344	0.468	0.353	0.297
	(0.38)	(0.61)	(0.84)	(0.54)	(0.86)
Dummy = 1 if MNC presence measured in employment	1.510^{*}	$1.436^{\dagger}$	2.216^{**}	1.808^{*}	1.411^{*}
	(2.23)	(1.77)	(2.94)	(2.42)	(2.45)
Dummy = 1 if MNC presence measured in assets	0.329	0.553	1.036	0.577	0.695
	(0.47)	(0.73)	(1.42)	(0.74)	(1.42)
Dummy = 1 if MNC presence measured in output	1.159	1.148	1.856^{*}	1.505	0.841
	(1.39)	(1.40)	(2.25)	(1.60)	(1.62)
Constant	20.85	39.86	44.44	8.379	37.03
	(0.35)	(0.57)	(0.66)	(0.13)	(0.51)
Observations	87	97	97	87	97
R^2 (pseudo R^2 for median reg. and probit)	0.342	0.258	0.128	0.335	0.238

Notes: OLS and RE computed excluding outliers; heteroscedasticity robust (Huber-White sandwich est.) t statistics in parentheses

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

not the case of our study since the sample that we use is very diversified, but still, panel-data methods might deliver more “balanced” results.

Another approach is to restrict the meta-response variable to a binary one and employ the probit or logit models (for a related example, see Wooster & Diebel 2006). Therefore, we construct a dummy variable which equals to one when t statistic is positive, and zero otherwise. Moreover, we construct a similar dummy for significance: if the absolute value of t statistic reaches the 5% critical value, the dummy equals one, and zero otherwise. Both models are estimated with normal probability regression and the details can be found in Table 3.14 in Section 3.C. Although there are slight differences between the results of the probit model when the response variable is dummy for positiveness (specification 1 from Table 3.14) and our benchmark-case OLS, basically it tells the same story in terms of significances and polarities of estimates.

When the dummy for significance is used as the meta-response variable, the only significant meta-explanatory variables are number of degrees of freedom in the study, average year of study period, and cross-sectionality of data. Our results suggest that higher number of observations leads to more significant results (either positive or negative), which is something one would expect. Cross-sectional data bring more significant t-statistics. Moreover, the reported degree of significance seems to be declining over time—studies using newer data are more likely to find insignificant results.

The results of all methods of meta-regression are summarized in Table 3.2. We prefer random effects since this method accounts for dependencies within studies and between-study heterogeneity. The results of other estimators are, nevertheless, not qualitatively different, which suggests that the results are robust to the particular methodology used. There are three meta-explanatory variables which are robustly significant at the 5% level. Our results show that cross-sectional data, industry-level aggregation, and usage of share in employment as a proxy for foreign presence brings, in general, more positively significant outcomes than other specifications. It does not seem to matter, on the other hand, how the response variable is defined.

Table 3.3: Summary of conducted meta-regressions, old studies

Response variable: t statistic; dummy = 1 if positive (probit)	OLS	IRLS	Median reg.	RE	Probit
Logarithm of degrees of freedom	0.137	0.163	$0.379^{\dagger}$	0.137	−0.121
	(0.63)	(0.76)	(1.82)	(0.63)	(−0.90)
Average year of study period	0.0265	0.0185	−0.0291	0.0265	−0.0285
	(0.69)	(0.42)	(−0.63)	(0.69)	(−0.51)
Dummy = 1 if data are for developing country	0.804	0.654	−0.547	0.804	−0.0816
	(1.00)	(0.69)	(−0.63)	(1.00)	(−0.09)
Dummy = 1 if data are for transition country	3.018^{**}	2.931^{*}	3.444^{**}	3.018^{**}	0.984
	(2.84)	(2.60)	(3.49)	(2.84)	(1.18)
Dummy = 1 if data are cross-section	$1.382^{\dagger}$	1.326	2.167^{*}	$1.382^{\dagger}$	1.810^{**}
	(1.95)	(1.42)	(2.44)	(1.95)	(2.93)
Dummy = 1 if response variable is output growth	0.527	0.434	−0.0435	0.527	−0.161
	(0.93)	(0.60)	(−0.07)	(0.93)	(−0.27)
Dummy = 1 if data are industry-level	3.057^{**}	3.168^{*}	3.580^{**}	3.057^{**}	
	(3.29)	(2.54)	(3.08)	(3.29)	
Dummy = 1 if industry dummies are used	0.191	0.506	0.787	0.191	−0.763
	(0.16)	(0.47)	(0.84)	(0.16)	(−0.93)
Dummy = 1 if MNC presence measured in employment	2.397^{*}	2.308^{*}	$1.650^{\dagger}$	2.397^{*}	0.701
	(2.30)	(2.04)	(1.71)	(2.30)	(0.80)
Dummy = 1 if MNC presence measured in assets	0.225	0.288	0.0177	0.225	1.183
	(0.21)	(0.29)	(0.02)	(0.21)	(1.60)
Dummy = 1 if MNC presence measured in output	4.433^{**}	4.383^{**}	4.559^{**}	4.433^{**}	$1.718^{\dagger}$
	(3.67)	(3.46)	(3.78)	(3.67)	(1.66)
Constant	−57.65	−41.82	51.75	−57.65	56.18
	(−0.75)	(−0.48)	(0.56)	(−0.75)	(0.51)
Observations	42	46	46	42	46
R^2 (pseudo R^2 for median reg. and probit)	0.626	0.549	0.288	0.626	0.419

Notes: OLS and RE computed excluding outliers; heteroscedasticity robust (Huber-White sandwich est.) t statistics in parentheses
† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 3.4: Summary of conducted meta-regressions, new studies

Response variable: t statistic; dummy = 1 if positive (probit)	OLS	IRLS	Median reg.	RE	Probit
Logarithm of degrees of freedom	0.183	0.248	0.351	0.132	−0.00712
	(1.15)	(1.20)	(0.83)	(0.79)	(−0.06)
Average year of study period	−0.150	−0.256*	−0.277	−0.119	0.0000351
	(−1.26)	(−2.17)	(−0.92)	(−1.07)	(0.00)
Dummy = 1 if data are for developing country	−0.0703	0.900	1.874	−0.356	
	(−0.04)	(0.61)	(0.46)	(−0.23)	
Dummy = 1 if data are for transition country	1.092	1.881	3.141	0.751	0.234
	(0.56)	(1.12)	(0.69)	(0.39)	(0.37)
Dummy = 1 if data are cross-section	2.687*	3.213*	2.988	2.249*	0.931
	(2.72)	(2.68)	(1.07)	(2.25)	(1.04)
Dummy = 1 if response variable is output growth	1.153	1.615†	1.187	0.818	0.160
	(1.21)	(1.77)	(0.46)	(0.91)	(0.26)
Dummy = 1 if data are industry-level	3.438*	4.595**	5.020	3.199*	0.656
	(2.27)	(3.10)	(1.37)	(2.11)	(0.66)
Dummy = 1 if industry dummies are used	0.936	1.579	2.856	1.002	0.675
	(0.84)	(1.60)	(1.05)	(0.97)	(1.15)
Dummy = 1 if MNC presence measured in employment	2.046	2.299	3.765	2.108	2.283*
	(1.66)	(1.54)	(0.87)	(1.28)	(2.50)
Dummy = 1 if MNC presence measured in assets	0.651	1.118	1.537	0.757	1.400
	(0.62)	(0.79)	(0.37)	(0.49)	(1.49)
Dummy = 1 if MNC presence measured in output	0.396	1.057	0.768	0.687	0.794
	(0.33)	(0.79)	(0.20)	(0.39)	(0.97)
Constant	295.9	503.2*	544.2	234.6	−1.516
	(1.26)	(2.16)	(0.91)	(1.07)	(−0.01)
Observations	45	51	51	45	51
R^2 (pseudo R^2 for median reg. and probit)	0.314	0.348	0.099	0.302	0.208

Notes: OLS and RE computed excluding outliers; heteroscedasticity robust (Huber-White sandwich est.) t statistics in parentheses
† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

The significance of the cross-sectionality of the original data set confirms the findings of Görg & Strobl (2001), who claim that the bias could be caused by time invariant variables which are not identified by the explanatory variables in cross-sectional spillover studies. Panel data methods can, on the other hand, uncover these effects, and thus are more reliable. Cross-sectional studies, especially in combination with industry-level data, can thus cause the causality problem—foreign investors may seek efficient and more productive industries for their investments, thus researchers would report a positive spillover effect, even if the particular industry had had high productivity long before MNCs entered it. On the other hand, Proença *et al.* (2006) argue that the classical panel data methods of spillover estimation may generate *downward* bias and they recommend using the extended generalized method of moments.

Contrary to Görg & Strobl (2001), we also find the level of aggregation and usage of share in employment as a proxy for foreign presence significant. Concerning the former, industry-level aggregation over heterogeneous firms may generally lead to biased results (Görg & Greenaway 2004), since it does not cope with firm-specific effects that can be correlated with foreign presence. Concerning the latter, employment intensive foreign investments could generate larger spillovers through the labor turnover channel, contrary to the sales intensive foreign investors who may, on the other hand, be more involved in the competition effect which has ambiguous impacts on host-country firms (Meyer & Sinani 2005). This could explain the significant coefficient that was obtained for the variable *EMPL* and might suggest that using a share of employment as a proxy for foreign presence is not misspecification, however, the definition of proxy for foreign presence deserves attention. Researchers should always check their outcomes on various definitions of proxies and try to explain possible different outcomes.[5]

It is also evident that the dominant specification of spillovers' testing has been changing over time. Since the first researchers followed the pioneering work of Caves

[5]There is a general problem connected with defining "foreign presence". As Castellani & Zanfei (2007) show, the common approach can cause downward bias in spillover estimates, since it assumes that changes of the same proportion in aggregate and foreign activities within an industry do not affect the response variable, whilst the contrary can be the case in reality.

(1974) and used cross-sectional data and industry-level aggregation, a little had changed before Haddad & Harrison (1993) published their study on Morocco, where they—using firm-level panel data—found evidence of negative horizontal spillovers due to the competition effect. Nevertheless, not many researchers used panel data again till 1999, where the other highly influential work Aitken & Harrison (1999) was published. After that, panel-data and firm-level analysis has become more frequent and has been almost unambiguously dominating the literature since 2003, leaving cross-sectional and industry-level methods mostly for countries where detailed data are not easily accessible, e.g., China. Because our results suggest that the (non)presence of time dimension in the data is one of the crucial aspects of the study design, we decide to split the sample into two halves (studies published before 2003, and *vice versa*), and employ the Chow test to check whether it was appropriate to pool the data together in the first place. The Chow test is significant only at the 23% level, thus the data were probably pooled correctly. Still, it might be benefitial to estimate the model separately for the two time periods.[6]

The results of meta-regressions for older studies are reported in Table 3.3. In the case of probit, the dummy for industry-level data had to be omitted since otherwise the probit model would not have converged.[7] The Breusch-Pagan Lagrange multiplier test is significant at the 10% level and the random-effects model is preferred to plain OLS. Similarly as for the pooled sample of all studies, it seems to matter whether data are cross-sectional, aggregated on the industry level, and whether the share of foreign presence is measured in employment. Contrary to the pooled sample, however, also the fact whether data for transition countries are used and whether foreign presence is measured as share in output are significant. In the older studies, firms in transition countries are more likely to benefit from horizontal FDI spillovers.

Results for newer studies can be found in Table 3.4. In the case of probit, one

[6] We also ran Chow test for equality of regression coefficients for developing and other countries subsamples, respectively. The null hypothesis was not rejected. When we estimated the model for both subsamples separately, the differences were not qualitatively important and thus are not reported.

[7] It does not mean, though, that *INDUSTRY* would be insignificant. Conversely, it predicts a perfect fit—industry-level aggregation always brings positive values of t statistics for spillovers in older studies.

dummy (developing country) had to be dropped so as for the model to converge. The Breusch-Pagan test is not significant at any reasonable level, thus we put more weight on plain OLS. Estimated dependencies are much less apparent now than for the older studies. It is again important whether data are cross-sectional and what the level of aggregation is, but no other meta-explanatory variable is significant in more than only one specification of Table 3.4. Thus it appears that the pattern, having basically still the same shape, is getting weaker over time. This would suggest that, at least recently, researchers have been aware of this dependency of results on the study design and have begun to employ more balanced approaches, maybe even to compensate for the "expected" results. Indeed, the empirical literature has been diverging a lot since the work of Görg & Strobl (2001) was published. A significant number of new studies test both for intra-industry and inter-industry spillovers, authors check multiple methodologies and compare the results. Nevertheless, there are still simple cross-sectional and/or industry-level studies, results of which can mostly be predicted *ex ante*.

3.6 Publication Bias

Stanley (2001) highlights the "file drawer" problem that occurs when researchers tend to publish only or mostly the studies that are able to demonstrate significant results or are consistent with the predominant theory because these are more likely to be accepted for publication in academic journals. It has been shown, e.g., by Card & Krueger (1995) that the "file drawer" problem can be extremely significant in economic publishing. In the concrete, for the literature on minimum wages and employment they find vast evidence of a publication bias. The same phenomena was detected by Görg & Strobl (2001) in the spillover literature and both subsequent meta-analyses (Meyer & Sinani 2005; Wooster & Diebel 2006) report similar results using the same methodology.

First, we employ the test advocated by Card & Krueger (1995) and applied by Görg & Strobl (2001) to the spillover literature. The set-up is illustrated in (3.5)—we

Table 3.5: Test of publication bias, OLS

Response variable: absolute value of t statistic	(1)	(2)	(3)	(4)
Logarithm of square root of degrees of freedom	1.165^{**}	1.161^{**}	0.865^{*}	1.163^{**}
	(2.95)	(2.71)	(2.33)	(2.70)
Average year of study period	−0.0235	−0.0168		
	(−0.59)	(−0.46)		
Dummy = 1 if data are for developing country	−0.564	−0.600		
	(−0.31)	(−0.43)		
Dummy = 1 if data are for transition country	−0.976	−0.196		−0.463
	(−0.38)	(−0.14)		(−0.27)
Dummy = 1 if data are cross-section	$3.419^{\dagger}$	2.605		3.403^{*}
	(1.98)	(1.63)		(2.07)
Dummy = 1 if response variable is output growth	1.620		1.220	1.684
	(1.26)		(0.96)	(1.50)
Dummy = 1 if data are industry-level	−0.515		0.597	−0.670
	(−0.42)		(0.67)	(−0.57)
Dummy = 1 if industry dummies are used	0.559		−0.534	
	(0.58)		(−0.39)	
Dummy = 1 if MNC presence measured in employment	−0.809		0.0417	−0.498
	(−0.29)		(0.02)	(−0.28)
Dummy = 1 if MNC presence measured in assets	0.0104		0.388	
	(0.01)		(0.21)	
Dummy = 1 if MNC presence measured in output	−1.532		−0.428	−1.415
	(−0.90)		(−0.28)	(−1.14)
Constant	45.06	31.81	−0.517	−2.177
	(0.56)	(0.44)	(−0.31)	(−1.09)
Observations	97	97	97	97
R^2	0.127	0.083	0.050	0.123
t $(H_0 : \beta = 1)$	0.170	0.140	0.130	0.140

Notes: heteroscedasticity robust (Huber-White sandwich est.) t statistics in parentheses

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

regress the absolute value of t statistics reported by the k-th model on the natural logarithm of the square root of number of degrees of freedom in the k-th model, controlling also for all other meta-explanatory variables which were included in model (3.3):

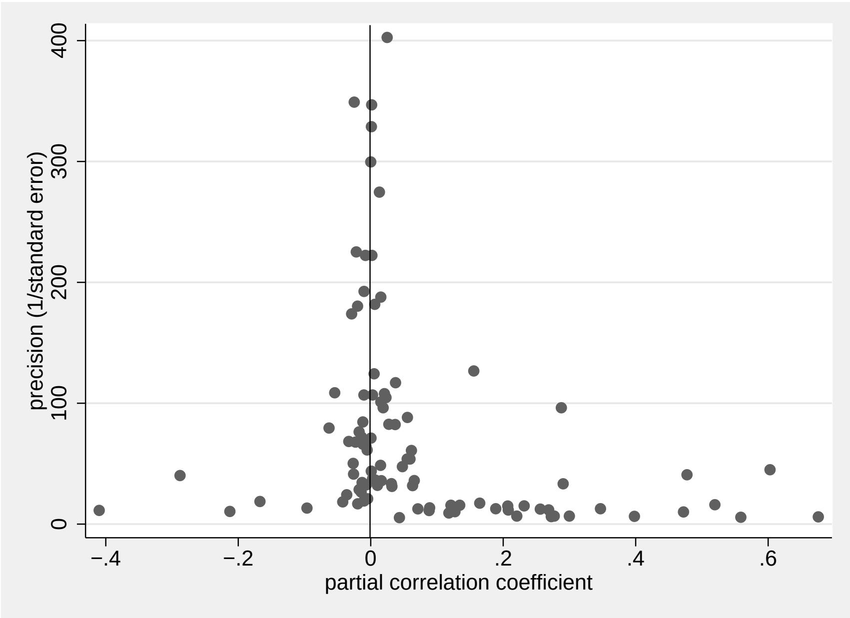
$$|t_k| = \alpha + \beta \log(\sqrt{M_k}) + \sum_{l=1}^{L-1} \gamma_l X_{kl} + \epsilon_k, \quad k = 1, 2, \ldots, K, \tag{3.5}$$

where M_k is the number of degrees of freefom in the k-th model. The crucial point of this test is the (non)significance and magnitude of the estimated parameter β. Under the null hypothesis of no publication bias, it should hold that $\beta = 1$. In other words, logarithm of square root of degrees of freedom should increase the final model's t statistic for foreign presence proportionally angle-wise 45 degrees.

Results of the publication bias test are reported in Table 3.5. It is a good sign that, under any specification, the estimate of β is significant at least at the 5% level and it is positive, which suggests that more degrees of freedom, *ceteris paribus*, increase the results' level of significance as it should be the case of unbiased literature. Estimated values of β are very close to 1 for all specifications. Testing the hypothesis $\beta = 1$ with a simple t test, we conclude that there is no sign of publication bias using this methodology (the corresponding test statistics are available in Table 3.5, as well).

However, the presented test for publication bias, employed by all previous meta-analyses of the spillover effect, was sharply criticized by Doucouliagos & Stanley (2009). They show that the methodology of Card & Krueger (1995) in fact confuses publication bias with testing for the underlying "true effect". An alternative way how to test for publication bias is to transform t statistics to partial correlation coefficients following Doucouliagos & Laroche (2009) and compute the corresponding standard errors. This conversion is necessary; regression coefficients cannot be used because they are not directly comparable. Partial correlation coefficients, similar to the simple correlation coefficients, show the statistical strength of the relationship between two variables—in this case, domestic firms' productivity and foreign presence in the sector. Now the most precise estimates of the underlying effect should lie close to the "true" partial correlation coefficient, and the estimates with lower precision should be more dispersed around this value. A natural measure of precision is an inverse value of the standard error; $1/se$. In the absence of publication bias, plotting precision against partial correlation coefficients should thus yield a symmetric inverted funnel (Stanley & Doucouliagos 2010a).

Figure 3.1: Funnel plot



Literature on horizontal FDI spillovers produces Figure 3.1. The funnel is not entirely symmetric, as the positive part is evidently heavier suggesting that negative estimates have lower chance of being published. Compared to other areas of empirical economics research, however, publication bias is moderate; see Doucouliagos & Stanley (2008). It is apparent that the estimates with highest precision are concentrated very close to zero, which suggests that the underlying effect beyond publication bias may be very small. Some researchers even argue that it might paradoxically be efficient to discard 90% of the data and draw inferences only from the 10% remaining observations with highest precision (Stanley & Doucouliagos 2010b). Since the funnel has only a single peak, we can also notice that our sample is relatively homogeneous. It is the nature of such visual tests, however, that they may be interpreted subjectively. Fortunately, a simple formalization of the "funnel asymmetry test" exists. It allows us to test the presence and magnitude of publication bias and the underlying effect of foreign presence on domestic firms' productivity.

In the absence of publication bias, the estimates should be randomly distributed around the true value (β) with no dependence on the standard error [β_1 in (3.6) should be zero]:

$$r_k = \beta + \beta_1 se_k + \epsilon_k, \tag{3.6}$$

where r denotes partial correlation coefficient estimated in the literature, se the corresponding standard error, and ϵ normal disturbance term. Because specification (3.6) is obviously heteroscedastic, the weighted least squares version is usually employed:

$$r_k/se_k = t_k = \beta/se_k + \beta_1 + \varphi_k. \tag{3.7}$$

Specification (3.7) can be also derived directly from the funnel plot by switching the axes and dividing partial correlation coefficients by standard errors to remove heteroscedasticity.

Table 3.6: Alternative test of publication bias

Response variable: t statistic	OLS	IRLS	RE
True effect ($1/se$)	-0.00328	-0.00654^{*}	-0.00189
	(-0.51)	(-2.15)	(-0.29)
Publication bias (constant)	1.904^{**}	1.234^{**}	1.822^{**}
	(2.83)	(3.50)	(2.63)
Observations	97	97	97

heteroscedasticity-robust t statistics in parentheses
† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Results of the funnel asymmetry test are reported in Table 3.6. Following Krassoi-Peach & Stanley (2009) and Doucouliagos & Stanley (2009), we use standard OLS as a benchmark and additionally employ robust and random effects versions of the regression (the random effects panel estimator accounts for heterogeneity between studies). All estimators yield similar results. The constant term is significant which suggests that publication bias is present in the spillover literature; its intensity (we obtain values from 1.2 to 1.9) can be classified as moderate following Stanley & Doucouliagos (2010a).[8] While majority of applied economics shows more severe

[8]Publication bias is not significant at the 5% level if only studies published after 2003 are included in the regression.

publication selection bias, the results contradict with those of the traditional test for publication bias introduced by Card & Krueger (1995) and first applied to the spillover literature by Görg & Strobl (2001). The new methodology allows us to estimate the "true" effect net of publication bias [β from (3.6)], which is also presented in Table 3.6. The true effect is (*negatively*) significant only in one specification—and even in this case it is very small and probably of little practical importance. We thus present evidence that the average reported spillover effect which is strongly positive (see Section 3.4) arises largely due to publication bias.

3.7 Conclusions

This paper presents a meta-analysis of the empirical literature on horizontal productivity spillovers from FDI. We gather a sample of 97 models from 67 studies published either in academic journals or as working papers. Using the vote-counting method, the spillover effect is not significant in general; employing the approach of Djankov & Murrell (2002), on the other hand, there is some evidence that positive spillovers from FDI might exist. Nevertheless, it is not the case of the narrower sample of studies that were published in the best economics journals or that use panel and firm-level data (and thus are more reliable)—their combined t statistics is insignificant almost in any case. Once publication selection bias is accounted for, the aggregated effect is insignificant, no matter what methodology is used. Therefore, the present authors argue that there is no general persuasive empirical evidence on the intra-industry spillovers. If there are any horizontal spillover effects, their signs and magnitudes vary from country to country and from industry to industry (we also find some evidence that employment-intensive foreign direct investment may generate relatively higher spillovers through labor turnover).

We further investigate which study aspects affect the reported significance and polarity of spillovers using a meta-regression analysis which was elaborated by Stanley & Jarrell (1989). Nevertheless, we use not only the standard ordinary least squares meta-regression (like Görg & Strobl 2001) but we also employ robust methods (iteratively re-weighted least squares and median regression) as well as pseudo-panel data methods (Meyer & Sinani 2005) and probability models (Wooster & Diebel 2006). Subject to several sensitivity checks we find that, in general, study results are predictably affected by its design, namely by the usage of cross-sectional or panel data, industry- or firm-level aggregation, and specification of the proxy of foreign presence in the industry. Our results suggest that cross-sectional studies tend to report excessively high spillovers, as well as models with industry-level aggregation and employment as a proxy for foreign presence do. However, this pattern appears to become weaker over time, suggesting that newer studies may suffer from such a bias less.

Following Card & Krueger (1995), we test for publication bias in the spillover literature. Contrary to Görg & Strobl (2001), we do not find evidence of publication bias employing this methodology. When the preferred funnel asymmetry test (Doucouliagos & Stanley 2009; Doucouliagos & Laroche 2009) is used, however, moderate publication bias is identified in the literature.

Many man-hours of economics and business researchers all over the world have been devoted to investigate horizontal spillovers from foreign direct investment. Is it "much ado about nothing" as Görg & Greenaway (2004) suggest in the title of their article? While the spillover effect is probably heterogeneous across different countries and industries, the worrying issue is that the results are *systematically* dependent on the chosen methodology. In other words, the researcher can influence her results *ex ante* by simply choosing a particular methodology. A strong consensus has formed in the international research community that firm-level panel data are the appropriate tool to test the presence of spillovers from foreign direct investment. For many countries, however, such detailed data are often difficult to construct, and cross-sectional studies are still being published. The outcome of such studies is predictable to a large extent. The pattern, however, does not concern only the nature of the data. Contrary to Görg & Strobl (2001), our meta-regression analysis shows that the definition of the proxy for foreign presence is important as well and can also bring predictable results. Unfortunately, many studies do not report sensitivity analysis with respect to the definition of foreign presence. When they do, as for instance Geršl (2008), they often find that the spillover effect is not robust. Such pattern of predictability is widespread in economics research. Indeed, Stanley (2001) shows how one of his older meta-regression analyses on the union wage premium (Jarrell & Stanley 1990), coincidentally published in the same issue as a new empirical study on the topic, precisely estimated the results of that study once its characteristics were plugged into the meta-regression. It is natural that heterogeneous research brings heterogeneous results. Researchers should be aware of the predictability pattern, best identified by meta-regression analyses, and report thorough robustness checks.

The other problematic issue of the spillover literature is publication bias. While the identified magnitude of bias is not extreme, it is high enough to produce statistically significant results where there are probably none in reality. In a meta-meta analysis (meta-regression analysis of meta-analyses), Doucouliagos & Stanley (2008) show that theory competition is crucial for the resulting empirical publication bias in the particular field. For example, the predominant (neoclassical) theory predicts the effect of raising minimum wage on employment to be negative. Doucouliagos & Stanley (2009) illustrate how it is harder for positive estimates to be published causing the naive average taken from the literature to be biased toward the negative values. Thus, in this respect, it is beneficial for each empirical field to have competitive theoretical background. Concerning the spillover literature, theory competition has increased considerably during the last two decades. Negative results became accepted without much doubt. It is thus possible that publication bias will become less problematic in the coming years. Indeed, funnel asymmetry test executed on the subsample of new studies is not significant at the 5% level indicating no formal evidence for publication selection among such studies.

Future research should focus on the inter-industry spillovers since they seem to be more promising, the number of empirical works in this field is growing and will soon be sufficient for a meta-regression analysis. Intra-industry productivity spillovers, on the other hand, appear to stay nonexistent or undetectable after correcting for publication bias.

References

Aitken, B. J. & A. E. Harrison (1999): "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela." *American Economic Review* **89(3)**: pp. 605–618.

Aslanoglu, E. (2000): "Spillovers effect of foreign direct investment on Turkish manufacturing industry." *Journal of International Development* **12**: pp. 1111–1130.

Audretsch, D. B. (1998): "Agglomeration and the Location of Innovative Activity." *Oxford Review of Economic Policy* **14(2)**: pp. 18–29.

Barrios, S. (2000): "Foreign Direct Investment Productivity Spillovers. Evidence From the Spanish Experience (1990-1994)." *Working Papers 2000-19*, FEDEA, Madrid.

Barrios, S., S. Dimelis, H. Louri, & E. Strobl (2002): "Efficiency Spillovers from Foreign Direct Investment in the EU Periphery: A Comparative Study of Greece, Ireland and Spain." *Working Papers 2002-02*, FEDEA, Madrid.

Barrios, S. & E. Strobl (2002): "Foreign direct investment and productivity spillovers: Evidence from the Spanish experience." *Review of World Economics* **138(3)**: p. 459–481.

Blalock, G. & P. J. Gertler (2005): "Welfare Gains from Foreign Direct Investment through Technology Transfer to Local Suppliers." *Working paper*, University of California, Berkeley.

Blomström, M. (1986): "Foreign Investment and Productive Efficiency: The Case of Mexico." *Journal of Industrial Economics* **35(1)**: pp. 97–110.

Blomström, M. & A. Kokko (1996): "Multinational Corporations and Spillovers." *CEPR Discussion Papers 1365*, Center for Economic Policy Research.

Blomström, M. & A. Kokko (2003): "The Economics of Foreign Direct Investment Incentives." *NBER Working Papers 9489*, National Bureau of Economic Research, Inc.

Blomström, M., A. Kokko, & M. Zejan (1994): "Host country competition, labor skills, and technology transfer by multinationals." *Review of World Economics* **130(3)**: pp. 521–533.

Blomström, M. & H. Persson (1983): "Foreign investment and spillover efficiency in an underdeveloped economy: Evidence from the Mexican manufacturing industry." *World Development* **11(6)**: pp. 493–501.

Blomström, M. & F. Sjöholm (1999): "Technology transfer and spillovers: Does local participation with multinationals matter?" *European Economic Review* **43(4-6)**: pp. 915–923.

Blomström, M. & E. N. Wolff (1994): "Multinational Corporations and Productivity Convergence in Mexico." In W. J. Baumol, R. R. Nelson, & E. N. Wolff (editors), "Convergence of Productivity: Cross National Studies and Historical Evidence," pp. 263–283. Oxford: Oxford University Press.

Bosco, M. G. (2001): "Does FDI contribute to technological spillovers and growth? A panel data analysis of Hungarian firms." *Transnational Corporations* **10(1)**: pp. 43–67.

Bouoiyour, J. (2003): "Labour Productivity, Technological Gap and Spillovers: Evidence From Moroccan Manufacturing Industries." *Working papers*, WPCATT, University of Pau.

Buckley, P. J., C. Wang, & J. Clegg (2002): "The impact of inward FDI in the performance of Chinese manufacturing firms." *Journal of International Business Studies* **33(4)**: pp. 637–655.

Buckley, P. J., C. Wang, & J. Clegg (2007): "The impact of foreign ownership, local ownership and industry characteristics on spillover benefits from foreign direct investment in China." *International Business Review* **16(2)**: pp. 142–158.

Bwalya, S. M. (2006): "Foreign direct investment and technology spillovers: Evidence from panel data analysis of manufacturing firms in Zambia." *Journal of Development Economics* **81(2)**: pp. 514–526.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review* **85(2)**: pp. 238–43.

Castellani, D. & A. Zanfei (2007): "Multinational companies and productivity spillovers: is there a specification error?" *Applied Economics Letters* **14(14)**: pp. 1047–1051.

Caves, R. E. (1974): "Multinational Firms, Competition, and Productivity in Host-Country Markets." *Economica* **41(162)**: pp. 176–93.

Chuang, Y. & C. Lin (1999): "Foreign direct investment, R&D and spillover efficiency: evidence from Taiwan's manufacturing firms." *Journal of Development Studies* **35**: pp. 117–137.

Corden, W. M. (1974): *Trade Policy and Economic Welfare.* London: Oxford University Press.

Crespo, N. & M. P. Fontoura (2007): "Determinant Factors of FDI Spillovers - What Do We Really Know?" *World Development* **35(3)**: pp. 410–425.

Damijan, J. P., B. Majcen, M. Rojec, & M. Knell (2001): "The role of FDI, R&D accumulation, and trade in transferring technology to transmission countries: evidence from firm panel data for eight transition countries." *Working Papers 2001/10*, Institute for Economic Research, University of Ljubljana.

Djankov, S. & B. M. Hoekman (2000): "Foreign Investment and Productivity Growth in Czech Enterprises." *World Bank Economic Review* **14(1)**: pp. 49–64.

Djankov, S. & P. Murrell (2002): "Enterprise Restructuring in Transition: A Quantitative Survey." *Journal of Economic Literature* **40(3)**: pp. 739–792.

Doucouliagos, H. & P. Laroche (2009): "Unions and Profits: A Meta-Regression Analysis." *Industrial Relations* **48(1)**: pp. 146–184.

Doucouliagos, H. & T. Stanley (2008): "Theory Competition and Selectivity: Are All Economic Facts Greatly Exaggerated?" *Economics Series 06*, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Doucouliagos, H. & T. D. Stanley (2009): "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* **47(2)**: pp. 406–428.

Driffield, N. (2001): "The Impact of Domestic Productivity of Inward Investment in the UK." *Manchester School* **69(1)**: pp. 103–19.

Fidrmuc, J. & I. Korhonen (2006): "Meta-analysis of the business cycle correlation between the euro area and the CEECs." *Journal of Comparative Economics* **34(3)**: pp. 518–537.

Flôres, R. G., M. P. Fontoura, & R. G. Santos (2000): "Foreign direct investment spillovers: what can we learn from Portuguese data?" *Economics Working Papers (Ensaios Economicos da EPGE) 4/2000*, Graduate School of Economics, Getulio Vargas Foundation (Brazil).

Fosfuri, A., M. Motta, & T. Ronde (2001): "Foreign direct investment and spillovers through workers' mobility." *Journal of International Economics* **53(1)**: p. 205–222.

Gallet, C. (2007): "A Comparative Analysis of the Demand for Higher Education: Results from a Meta-analysis of Elasticities." *Economics Bulletin* **9(7)**: pp. 1–14.

Geršl, A. (2008): "Productivity, Export Performance, and Financing of the Czech Corporate Sector: The Effects of Foreign Direct Investment." *Czech Journal of Economics and Finance* **58**: pp. 232–247.

Geršl, A. & M. Hlaváček (2007): "Foreign Direct Investment, Corporate Finance, and the Life Cycle of Investment." *Czech Journal of Economics and Finance* **57**: pp. 448–464.

Girma, S. (2003): "Absorptive capacity and productivity spillovers from FDI: A threshold regression analysis." *Working paper series 25*, European Economy Group.

Girma, S., D. Greenaway, & K. Wakelin (2001): "Who Benefits from Foreign Direct Investment in the UK?" *Scottish Journal of Political Economy* **48(2)**: pp.

119–33.

Girma, S. & K. Wakelin (2007): "Local productivity spillovers from foreign direct investment in the U.K. electronics industry." *Regional Science and Urban Economics* **37**: pp. 399–412.

Globerman, S. (1979): "Foreign Direct Investment and 'Spillover' Efficiency Benefits in Canadian Manufacturing Industries." *Canadian Journal of Economics* **12(1)**: pp. 42–56.

Görg, H. & D. Greenaway (2004): "Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment?" *World Bank Research Observer* **19(2)**: pp. 171–197.

Görg, H. & E. Strobl (2001): "Multinational Companies and Productivity Spillovers: A Meta-analysis." *Economic Journal* **111(475)**: pp. F723–39.

Görg, H. & E. Strobl (2004): "Foreign direct investment and local economic development: Beyond productivity spillovers." *GEP Research Paper 2004/11*, The University of Nottingham.

Griffith, R., S. Redding, & J. V. Reenen (2003): "R&D and Absorptive Capacity: Theory and Empirical Evidence." *Scandinavian Journal of Economics* **105(1)**: pp. 99–118.

Haddad, M. & A. Harrison (1993): "Are there positive spillovers from direct foreign investment? Evidence from panel data for Morocco." *Journal of Development Economics* **42(1)**: pp. 51–74.

Hamilton, L. C. (2006): *Statistics with STATA*. Duxbury Press.

Haskel, J. E., S. C. Pereira, & M. J. Slaughter (2004): "Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?" *Working paper series*, Tuck School of Business at Darmouth.

Havránek, T. (2010): "Rose effect and the euro: Is the magic gone?" *Review of World Economics* **(forthcoming)**.

Hedges, L. V. (1992): "Meta-analysis." *Journal of Educational and Behavioral*

Statistics **17(4)**: pp. 279–296.

Hedges, L. V. & I. Olkin (1985): *Statistical Methods of Meta-Analysis.* Orlando: Academic Press.

Hermes, N. & R. Lensink (2003): "Foreign direct investment, financial development and economic growth." *The Journal of Development Studies* **40(1)**: pp. 142–163.

Huber, P. J. (1967): "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions." In "Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability," volume 1, p. 221–223. Berkeley, CA: University of California Press.

Hunter, J. & F. Schmidt (1990): *Methods of Meta-analysis: Correcting Error and Bias in Research Findings.* Newbury Park (CA): Sage Publications, 1st. edition.

Imbriani, C. & F. Reganati (1999): "Productivity spillovers and regional differences: some evidence on the Italian manufacturing sector." *Discussion Paper Series 48*, Universita degli Studi di Salerno.

Jarrell, S. B. & T. D. Stanley (1990): "A meta-analysis of the union-nonunion wage gap." *Industrial and Labor Relations Review* **44(1)**: pp. 54–67.

Javorcik, B. S. (2004): "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages." *American Economic Review* **94(3)**: pp. 605–627.

Javorcik, B. S. & M. Spatareanu (2008): "To share or not to share: Does local participation matter for spillovers from foreign direct investment?" *Journal of Development Economics* **85(1-2)**: pp. 194–217.

Jordaan, J. A. (2005): "Determinants of FDI-induced externalities: New empirical evidence for Mexican manufacturing industries." *World Development* **33(12)**: pp. 2103–2118.

Kalaitzidakis, P., T. P. Mamuneas, & T. Stengos (2003): "Rankings of Academic Journals and Institutions in Economics." *Journal of the European Economic Association* **1(6)**: pp. 1346–1366.

Kathuria, V. (2000): "Productivity spillovers from technology transfer to Indian manufacturing firms." *Journal of International Development* **12**: pp. 343–369.

Kathuria, V. (2002): "Liberalisation, FDI, and productivity spillovers—an analysis of Indian manufacturing firms." *Oxford Economic Papers* **54(4)**: pp. 688–718.

Keller, W. & S. R. Yeaple (2003): "Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States." *NBER Working Papers 9504*, National Bureau of Economic Research, Inc.

Khawar, M. (2003): "Productivity and foreign direct investment—evidence from Mexico." *Journal of Economic Studies* **30(1)**: pp. 66–76.

Kinoshita, Y. (2000): "R&D and Technology Spillovers via FDI: Innovation and Absorptive Capacity." *Working Papers Series 349*, William Davidson Institute at the University of Michigan Stephen M. Ross Business School.

Kohpaiboon, A. (2006): "Foreign direct investment and technology spillover: A cross-industry analysis of Thai manufacturing." *World Development* **34(3)**: pp. 541–556.

Kokko, A. (1994): "Technology, market characteristics, and spillovers." *Journal of Development Economics* **43(2)**: pp. 279–293.

Kokko, A. (1996): "Productivity spillovers from competition between local firms and foreign affiliates." *Journal of International Development* **8**: pp. 517–530.

Kokko, A. & M. Blomström (1995): "Policies to encourage inflows of technology through foreign multinationals." *World Development* **23(3)**: pp. 459–468.

Kokko, A., M. Zejan, & R. Tansini (1996): "Local technological capability and productivity spillovers from FDI in the Uruguayan manufacturing sector." *Journal of Development Studies* **32**: pp. 602–611.

Kokko, A., M. Zejan, & R. Tansini (2001): "Trade regimes and spillover effects of FDI: Evidence from Uruguay." *Review of World Economics* **137(1)**: pp. 124–149.

Konings, J. (2000): "The Effect of Direct Foreign Investment on Domestic Firms: Evidence from Firm Level Panel Data in Emerging Economies." *Working Paper*

Series 344, William Davidson Institute.

Krassoi-Peach, E. & T. Stanley (2009): "Efficiency Wages, Productivity and Simultaneity: A Meta-Regression Analysis." *Journal of Labor Research* **30(3)**: pp. 262–268.

Lee, J.-Y. & E. Mansfield (1996): "Intellectual Property Protection and U.S. Foreign Direct Investment." *The Review of Economics and Statistics* **78(2)**: pp. 181–86.

Li, H., T. Grijalva, & R. P. Berrens (2007): "Economic growth and environmental quality: a meta-analysis of environmental Kuznets curve studies." *Economics Bulletin* **17(5)**: pp. 1–11.

Lipsey, R. & F. Sjöholm (2004): "FDI and wage spillovers in Indonesian manufacturing." *Review of World Economics (Weltwirtschaftliches Archiv)* **140(2)**: pp. 321–332.

Liu, X., D. Parker, K. Vaiyda, & Y. Wei (2001): "The impact of foreign direct investment on labour productivity in the Chinese electronics industry." *International Business Review* **10**: p. 421–439.

Liu, X., P. Siler, C. Wang, & Y. Wei (2000): "Productivity Spillovers From Foreign Direct Investment: Evidence From UK Industry Level Panel Data." *Journal of International Business Studies* **31(3)**: pp. 407–425.

Liu, X. & C. Wang (2003): "Does foreign direct investment facilitate technological progress? Evidence from Chinese industries." *Research Policy* **32(6)**: pp. 945–953.

Liu, Z. (2002): "Foreign Direct Investment and Technology Spillover: Evidence from China." *Journal of Comparative Economics* **30(3)**: pp. 579–602.

Liu, Z. (2008): "Foreign direct investment and technology spillovers: Theory and evidence." *Journal of Development Economics* **85(1-2)**: pp. 176–193.

Lutz, S. & O. Talavera (2004): "Do Ukrainian Firms Benefit from FDI?" *Economics of Planning* **37(2)**: pp. 77–98.

MacDougall, G. D. A. (1960): "The Benefits and Costs of Private Investment from

Abroad: A Theoretical Approach." *Economic Record* **36**: pp. 13–35.

Marin, A. & M. Bell (2004): "Technology spillovers from foreign direct investment: the active role of MNC subsidiaries in Argentina in the 1990s." *SPRU Electronic Working Paper Series 118*, Science and Technology Policy Research, UK.

Merlevede, B. & K. Schoors (2006): "FDI and the Consequences Towards more complete capture of spillover effects." *Working Papers 06/372*, Ghent University, Faculty of Economics and Business Administration.

Meyer, K. E. & E. Sinani (2005): "Spillovers from Foreign Direct Investment: A Meta Analysis." *Working Paper Series 59*, Center for East European Studies, Copenhagen Business School.

Murakami, Y. (2007): "Technology spillover from foreign-owned firms in Japanese manufacturing industry." *Journal of Asian Economics* **18(2)**: pp. 284–293.

Narula, R. & A. Marin (2005): "Exploring the relationship between direct and indirect spillovers from FDI in Argentina." *Research Memoranda 024*, Maastricht Economic Research Institute on Innovation and Technology.

Nguyen, P. L. (2008): "Productivity Spillovers from Foreign Direct Investment: Evidence from Vietnamese Firm Data." *Working paper series*, School of Commerce, University of South Australia.

Pack, H. & K. Saggi (1997): "Inflows of Foreign Technology and Indigenous Technological Development." *Review of Development Economics* **1(1)**: pp. 81–98.

Peri, G. & D. Urban (2006): "Catching-up to foreign technology? Evidence on the "Veblen-Gerschenkron" effect of foreign investments." *Regional Science and Urban Economics* **36(1)**: pp. 72–98.

Proença, I., P. Fontoura, & N. Crespo (2006): "Productivity Spillovers from Multinational Corporations in Portugal: Vulnerability to Deficient Estimation." *Applied Econometrics and International Development* **6(1)**: pp. 87–98.

Ran, J., J. P. Voon, & G. Li (2007): "How does FDI affect China? Evidence from industries and provinces." *Journal of Comparative Economics* **35(4)**: pp. 774–799.

RATTSØ, J. & H. E. STOKKE (2003): "Learning and foreign technology spillovers in Thailand: Empirical evidence on productivity dynamics." *Working paper series*, Department of Economics, Norwegian University of Science and Technology.

RUANE, F. & A. UGUR (2003): "Foreign direct investment and productivity spillovers in Irish manufacturing industry: evidence from plant level panel data." *Working paper series*, Department of Economics, Trinity College Dublin.

SADIK, A. T. & A. A. BOLBOL (2001): "Capital Flows, FDI, and Technology Spillovers: Evidence from Arab Countries." *World Development* **29(12)**: pp. 2111–2125.

SASIDHARAN, S. & A. RAMANATHAN (2007): "Foreign Direct Investment and spillovers: evidence from Indian manufacturing." *The International Journal of Trade and Global Markets* **1(1)**: p. 5–22.

SCHOORS, K. & B. V. D. TOL (2002): "Foreign direct investment spillovers within and between sectors: Evidence from Hungarian data." *Working Papers 02/157*, Ghent University, Faculty of Economics and Business Administration.

SGARD, J. (2001): "Direct Foreign Investments And Productivity Growth In Hungarian Firms, 1992-1999." *Working Papers Series 425*, William Davidson Institute at the University of Michigan Stephen M. Ross Business School.

SINANI, E. & K. E. MEYER (2004): "Spillovers of technology transfer from FDI: the case of Estonia." *Journal of Comparative Economics* **32(3)**: pp. 445–466.

SJÖHOLM, F. (1999a): "Productivity Growth in Indonesia: The Role of Regional Characteristics and Direct Foreign Investment." *Economic Development and Cultural Change* **47(3)**: pp. 559–84.

SJÖHOLM, F. (1999b): "Technology Gap, Competition and Spillovers from Direct Foreign Investment: Evidence from Establishment Data." *Journal of Development Studies* **36**: pp. 53–73.

STANLEY, T. & H. DOUCOULIAGOS (2010a): "A simple graph that reveals much ado about research." *Journal of Economic Surveys* **24(1)**: pp. 170–191.

STANLEY, T. & H. DOUCOULIAGOS (2010b): "Could it be better to discard 90% of the data? A statistical paradox." *The American Statistician* **64(1)**: pp. 70–77.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. & S. B. Jarrell (1989): "Meta-Regression Analysis: A Quantitative Method of Literature Surveys." *Journal of Economic Surveys* **3(2)**: pp. 161–70.

Takii, S. (2005): "Productivity spillovers and characteristics of foreign multinational plants in Indonesian manufacturing 1990–1995." *Journal of Development Economics* **76(2)**: pp. 521–542.

Thuy, L. T. (2005): "Technological Spillovers from Foreign Direct Investment: the case of Vietnam." *Working paper*, Graduate School of Economics, University of Tokio.

Torlak, E. (2004): "Foreign Direct Investment, Technology Transfer and Productivity Growth in Transition Countries: Empirical Evidence from Panel Data." *CEGE Discussion Papers 26*, Center for European, Governance and Economic Development Research, University of Goettingen (Germany).

Vahter, P. (2004): "The Effect Of Foreign Direct Investment On Labour Productivity: Evidence From Estonia And Slovenia." *Working Paper Series 32*, Faculty of Economics and Business Administration, University of Tartu.

Víšek, J. A. (1997): *Econometrics I.* Prague: Charles University.

Wang, J.-Y. & M. Blomstrom (1992): "Foreign investment and technology transfer : A simple model." *European Economic Review* **36(1)**: pp. 137–155.

Wei, Y. & X. Liu (2003): "Productivity spillovers among OECD, diaspora and indigenous firms in Chinese manufacturing." *Working Papers 2003/008*, Lancaster University Management School, Economics Department.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* **48(4)**: pp. 817–38.

Wooster, R. B. & D. S. Diebel (2006): "Productivity Spillovers from Foreign Direct Investment in Developing Countries: A Meta-Regression Analysis." *Working* *paper*, Available at SSRN: http://ssrn.com/abstract=898400.

Xu, B. (2000): "Multinational enterprises, technology diffusion, and host country productivity growth." *Journal of Development Economics* **62(2)**: pp. 477–493.

Yudaeva, K., K. Konstantin, N. Melentieva, & N. Ponomareva (2000): "Does Foreign Ownership Matter? Russian Experience." *Working Papers w0005*, Center for Economic and Financial Research (CEFIR).

Zemplinerová, A. & M. Jarolím (2001): "Modes of FDI Entry and Firm Performance in Transition: The Czech Case." *Transnational Corporations* **10**: pp. 95–111.

Zhu, G. & K. Y. Tan (2000): "Foreign direct investment and labor productivity: New evidence from China as the Host." *Thunderbird International Business Review* **42(5)**: pp. 507–528.

3.A Data Description

On the following pages we provide a few tables illustrating the properties of studies used and the characteristics of regression variables.

Table 3.7: Studies used in the meta-analysis

Authors	Country	Years	Data	Level	Presence	Response	Result (5%)
Caves (1974)	Australia	1966	cs	industry	empl	prod	+
Globerman (1979)	Canada	1972	cs	industry	output	prod	?
Blomström & Persson (1983)	Mexico	1970	cs	industry	empl	prod	+
Blomström (1986)	Mexico	1970/75	cs	industry	empl	other	+
Haddad & Harrison (1993)	Morocco	1985-90	panel	secdum	assets	prod, growth	-
Blomström & Wolff (1994)	Mexico	1970/75	cs	industry	empl	growth	+
Kokko (1994)	Mexico	1970	cs	industry	empl	prod	+
Kokko (1996)	Mexico	1970	cs	industry	empl	prod	+
Kokko *et al.* (1996)	Uruguay	1970	cs	firm	output	prod	?
Aitken & Harrison (1999)	Venezuela	1976-89	panel	firm	assets	growth	-
Blomström & Sjöholm (1999)	Indonesia	1991	cs	secdum	output	prod	+
Chuang & Lin (1999)	Taiwan	1991	cs	firm	assets	growth	+
Imbriani & Reganati (1999)	Italy	1992	cs	industry	empl	prod	+
Sjöholm (1999b)	Indonesia	1980-91	cs	firm	output	prod, growth	+
Sjöholm (1999a)	Indonesia	1980-91	cs	firm	output	growth	+
Aslanoglu (2000)	Turkey	1993	cs	industry	sales	prod	?
Barrios (2000)	Spain	1990-94	panel	firm	output	prod	?
Djankov & Hoekman (2000)	Czech Republic	1993-96	panel	firm	assets	growth	-
Flôres *et al.* (2000)	Portugal	1992-95	panel	firm	output	prod	?
Kathuria (2000)	India	1976-89	panel	firm	sales	other	-
Kinoshita (2000)	Czech Republic	1995-98	panel	secdum	empl	growth	?
Konings (2000)	Trans. countries	1993-97	panel	firm	sales	growth	?
Liu *et al.* (2000)	UK	1991-95	panel	industry	empl	prod	+

Continued on next page

Studies used in the meta-analysis (continued)

Authors	**Country**	**Years**	**Data**	**Level**	**Presence**	**Response**	**Result (5%)**
Yudaeva *et al.* (2000)	Russia	1993-97	panel	secdum	output	prod	?
Bosco (2001)	Hungary	1993-97	panel	firm	sales	sales	?
Damijan *et al.* (2001)	Trans. countries	1994-98	panel	secdum	sales	growth	?
Driffield (2001)	UK	1989-92	cs	industry	sales	growth	?
Girma *et al.* (2001)	UK	1991-96	panel	firm	empl	prod, growth	?
Liu *et al.* (2001)	China	1996-97	cs	industry	assets	prod	?
Sgard (2001)	Hungary	1992-99	panel	firm	assets	prod	?
Zemplinerová & Jarolím (2001)	Czech Republic	1994-98	panel	secdum	assets	growth	?
Barrios *et al.* (2002)	Greece, Ireland, Spain	1992, 97	cs	firm	empl	prod	?
Buckley *et al.* (2002)	China	1995	cs	industry	assets, empl	prod	+
Kathuria (2002)	India	1990-96	panel	firm	sales	growth	-
Liu (2002)	China	1993-98	panel	secdum	assets	prod	?
Schoors & Tol (2002)	Hungary	1997-98	cs	firm	sales	prod	+
Bouoiyour (2003)	Morocco	1987-96	panel	industry	assets	prod	?
Khawar (2003)	Mexico	1990	cs	firm	assets	prod	?
Keller & Yeaple (2003)	USA	1987-96	panel	secdum	empl	growth	+
Liu & Wang (2003)	China	1995	cs	industry	assets	prod	+
Ruane & Ugur (2003)	Ireland	1991-98	panel	secdum	empl	growth	+
Wei & Liu (2003)	China	2000	cs	industry	assets	prod	+
Görg & Strobl (2004)	Ireland	1973-95	panel	firm	empl	growth	-
Haskel *et al.* (2004)	UK	1973-92	panel	firm	empl	growth	+
Javorcik (2004)	Lithuania	1996-00	panel	firm	assets	prod	+
Lutz & Talavera (2004)	Ukraine	1998-99	cs	secdum	assets	prod	?

Continued on next page

Studies used in the meta-analysis (continued)

Authors	**Country**	**Years**	**Data**	**Level**	**Presence**	**Response**	**Result (5%)**
Marin & Bell (2004)	Argentina	1992-96	panel	firm	empl	growth	?
Sinani & Meyer (2004)	Estonia	1994-99	panel	secdum	various	growth	+
Torlak (2004)	Trans. countries	1993-00	panel	secdum	output	prod	?
Vahter (2004)	Estonia, Slovenia	1994-01	panel	secdum	assets	prod	?
Blalock & Gertler (2005)	Indonesia	1988-96	panel	firm	output	prod	?
Jordaan (2005)	Mexico	1993	cs	firm	empl	prod	?
Narula & Marin (2005)	Argentina	92-96, 98-01	panel	secdum	empl	growth	?
Takii (2005)	Indonesia	1990-95	panel	firm	empl	prod	?
Thuy (2005)	Vietnam	1995-02	panel	industry	empl	prod	?
Bwalya (2006)	Zambia	1993-95	panel	firm	empl	growth	?
Kohpaiboon (2006)	Thailand	1996	cs	firm	output	prod	-
Merlevede & Schoors (2006)	Romania	1996-01	panel	firm	output	growth	?
Peri & Urban (2006)	Germany, Italy	1993-99	panel	secdum	empl	prod	+
Ran *et al.* (2007)	China	2001-03	panel	industry	assets	prod	?
Buckley *et al.* (2007)	China	2001	cs	industry	assets	prod	+
Girma & Wakelin (2007)	UK	1980-92	panel	firm	empl	prod	?
Murakami (2007)	Japan	1994-98	panel	secdum	empl	growth	?
Sasidharan & Ramanathan (2007)	India	1994-02	panel	secdum	output	growth	?
Javorcik & Spatareanu (2008)	Romania	1998-03	panel	firm	output	growth	?
Liu (2008)	China	1995-99	panel	secdum	assets	prod	-
Nguyen (2008)	Vietnam	2000-05	panel	secdum	output	prod	?

Notes: The column “Result” does not necessarily report researchers’ conclusion; the significance of spillover effect is based on simple average of specifications which were included to our analysis from the particular paper.

Table 3.8: Variable Characteristics

Variable	Definition	Summary stat.
	Response variable	
tstat	papers' t-statistics for foreign presence; meta-response variable	1.576 (5.65)
growth	= 1 if growth is response variable used in literature, = 0 if labor productivity	39
	Foreign Presence Measures	
empl	= 1 if MNC presence measured in employment, = 0 if otherwise (as output, assets, sales)	32
output	= 1 if MNC presence measured in output, = 0 if otherwise (as employment, assets, sales)	21
assets	= 1 if MNC presence measured in assets, = 0 if otherwise (as employment, output, sales)	25
	Data Specification	
cs	= 1 if data are cross-section, = 0 if panel data	32
industry	= 1 if data are industry-level, =0 if firm-level	22
secdum	= 1 if industry dummies are used, = 0 if otherwise	35
trans	= 1 if data are for transition country, = 0 if otherwise (developing, advanced)	34
devg	= 1 if data are for developing country, = 0 if otherwise (transition, advanced)	41
avgyr	Average year of study period	1992.286 (7.835)
ldf	Logarithm of degrees of freedom	7.377 (2.356)

Notes: For tstat, avgyr and ldf, the summary statistics is the mean with standard error in parenthesis, for all others it is the number of observations for which dummy variable equals 1.

3.B Multicollinearity

If we regress in turn all explanatory variables on the remaining explanatory variables and collect the coefficients of determination of the corresponding regressions, we obtain the linear redundancy statistics (see Table 3.9). The highest R-squared reaches 0.67, which is not excessive.

Table 3.9: Linear and non-linear relationships

Variable	Linear	Polynomial
Logarithm of degrees of freedom	0.457	0.497
Average year of study period	0.322	0.389
Dummy = 1 if data are for developing country	0.532	0.618
Dummy = 1 if data are for transition country	0.665	0.755
Dummy = 1 if data are cross-section	0.455	0.487
Dummy = 1 if response variable is output growth	0.279	0.330
Dummy = 1 if data are industry-level	0.547	0.699
Dummy = 1 if industry dummies are used	0.308	0.355
Dummy = 1 if MNC presence measured in employment	0.656	0.687
Dummy = 1 if MNC presence measured in assets	0.548	0.570
Dummy = 1 if MNC presence measured in output	0.562	0.595

An important thing—which is, nevertheless, usually omitted—is to test also for non-linear relationships between explanatory variables (Víšek 1997, p. 71). Such relationships cannot be discovered by standard correlation and redundancy analysis. Suppose for example that we obtain the following estimate of a regression model:

$$\widehat{Y}_i = X_{i1} + 2X_{i2}, \quad i = 1, 2, \ldots, N. \tag{3.8}$$

Assume also that there is a latent relationship which would give estimate $\widehat{X}_{i2} = 1 - 10X_{i1}^4$. If one obtains (3.8) and claims on the basis of it that X_{i1} has positive impact on Y_i, it is obviously not correct. This issue is even more problematic for studies which report polarities of some regression estimates as their key results—and this is the case of empirical works on productivity spillovers. A way how to (try to) discover such non-linear relationships is to use the Weierstrass Approximation

Theorem and estimate J following regressions:

$$X_{im} = \alpha + \sum_{j=1}^{J} \sum_{p=1}^{P} \beta_{jp} X_{ij}^{p} + \vartheta_i, \quad i = 1, \ldots, N, \quad m = 1, \ldots, J, \quad m \neq j, \tag{3.9}$$

where one must have $JP < N$ to leave a sufficient number of degrees of freedom for the regressions. We solved (3.9) with $J = 11$ and $P = 6$, the coefficients of determination are listed in Table 3.9. The highest increase in R-squared compared to simple linear redundancy was detected for variable $INDUSTRY$ and reached 0.15, which is not much taking into account that the new regression has 50 more explanatory variables. Therefore we can conclude that non-linear relationships do not represent a substantial problem for our analysis.

3.C Supplementary Tables

Additional material is available on the journal website.

Table 3.10: Table of correlation coefficients

	ldf	**avgyr**	**devg**	**trans**	**cs**	**growth**	**industry**	**secdum**	**empl**	**assets**	**output**
ldf	1										
avgyr	.289	1									
devg	-.235	-.245	1								
trans	.153	.383	-.629	1							
cs	-.423	-.431	.376	-.424	1						
growth	.185	.093	-.276	.279	-.352	1					
industry	-.602	-.310	.384	-.398	.510	-.344	1				
secdum	.272	.33	-.295	.438	-.436	.216	-.407	1			
empl	-.0665	-.274	-.0233	-.424	.0673	.006	.196	-.162	1		
assets	-.0259	.207	.212	.0117	.0377	-.147	.131	-.001	-.413	1	
output	.223	-.045	.0569	.0335	.0571	-.176	-.225	.0742	-.369	-.31	1

Table 3.11: Standard meta-regression, all studies

	OLS, including outliers				OLS, excluding outliers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ldf	0.340	0.317			0.0969	0.00201		
	(1.26)	(1.28)			(0.69)	(0.02)		
avgyr	0.0635	0.0457			−0.0119	−0.0301		
	(1.29)	(0.98)			(−0.40)	(−0.93)		
devg	−1.323	−1.972			−0.124	−0.461		
	(−0.65)	(−1.26)			(−0.23)	(−0.74)		
trans	1.080	0.338		2.173	0.805	−0.0612		0.824
	(0.40)	(0.21)		(1.17)	(0.99)	(−0.10)		(1.35)
cs	6.106**	5.185**		4.901**	2.023**	2.118**		1.910**
	(3.06)	(3.01)		(2.80)	(3.16)	(2.90)		(3.59)
growth	1.995		1.376	1.682	0.973†		0.839	0.923*
	(1.44)		(0.93)	(1.32)	(1.91)		(1.56)	(2.01)
industry	1.153		1.704	−0.535	1.851**		2.333**	1.547*
	(0.83)		(1.61)	(−0.33)	(2.85)		(4.19)	(2.63)
secdum	1.627		0.902		0.237		0.0251	
	(1.46)		(0.57)		(0.38)		(0.04)	
empl	2.376		1.878	1.988	1.510*		1.276*	1.365*
	(0.87)		(0.98)	(1.04)	(2.23)		(2.34)	(2.45)
assets	1.118		1.016		0.329		0.165	
	(0.55)		(0.46)		(0.47)		(0.23)	
output	1.019		1.563	0.501	1.159		1.390†	1.076†
	(0.55)		(0.86)	(0.37)	(1.39)		(1.95)	(1.69)
Constant	−132.1	−92.86	−0.909	−2.122	20.85	59.93	−1.191†	−1.846**
	(−1.35)	(−1.00)	(−0.49)	(−1.49)	(0.35)	(0.93)	(−1.82)	(−2.70)
Observations	97	97	97	97	87	87	87	87
R^2	0.185	0.131	0.031	0.133	0.342	0.222	0.232	0.331

Notes: ldf is a logarithm of degrees of freedom, *avgyr* the average year of study period, *devg* equals 1 if data are for developing country, *trans* equals 1 if data are for transition country, *cs* equals 1 if data are cross-sectional, *growth* equals 1 if response variable is output growth, *industry* equals 1 if data are industry-level, *secdum* equals 1 if industry dummies are used, *empl* equals 1 if MNC presence is measured in employment, *assets* for foreign presence measured in assets and *output* for foreign presence measured in output.

Table 3.12: Robust meta-regression, all studies

	IRLS				Median regression			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ldf	0.137	0.0387			0.100	0.110		
	(1.06)	(0.35)			(0.78)	(1.13)		
avgyr	−0.0216	−0.0356			−0.0239	−0.0250		
	(−0.62)	(−1.05)			(−0.71)	(−0.84)		
devg	0.0353	−0.0384			−0.0411	−0.183		
	(0.05)	(−0.06)			(−0.07)	(−0.34)		
trans	0.833	0.0195		0.784	1.068	−0.0503		$1.000^{\dagger}$
	(1.03)	(0.03)		(1.33)	(1.37)	(−0.09)		(1.67)
cs	1.876^{**}	1.965^{**}		1.821^{**}	2.363^{**}	2.835^{**}		2.579^{**}
	(2.91)	(3.25)		(3.09)	(3.70)	(5.21)		(4.34)
growth	0.880		0.757	0.849	0.839		0.160	0.360
	(1.64)		(1.38)	(1.66)	(1.57)		(0.27)	(0.68)
industry	1.884^{*}		2.299^{**}	1.552^{*}	0.770		2.120^{**}	0.481
	(2.37)		(3.45)	(2.28)	(1.03)		(2.95)	(0.70)
secdum	0.344		0.126		0.468		0.457	
	(0.61)		(0.23)		(0.84)		(0.80)	
empl	$1.436^{\dagger}$		$1.282^{\dagger}$	1.216^{*}	2.216^{**}		$1.240^{\dagger}$	1.593^{**}
	(1.77)		(1.85)	(2.14)	(2.94)		(1.67)	(2.78)
assets	0.553		0.442		1.036		0.483	
	(0.73)		(0.60)		(1.42)		(0.61)	
output	1.148		$1.334^{\dagger}$	$1.097^{\dagger}$	1.856^{*}		1.280	1.553^{*}
	(1.40)		(1.71)	(1.76)	(2.25)		(1.55)	(2.44)
Constant	39.86	70.61	−1.165	-1.679^{**}	44.44	49.06	−0.940	-1.553^{*}
	(0.57)	(1.05)	(−1.61)	(−2.81)	(0.66)	(0.83)	(−1.21)	(−2.56)
Observations	97	97	97	97	97	97	97	97
R^2	0.258	0.183	0.173	0.248	0.128	0.091	0.073	0.102

Notes: ldf is a logarithm of degrees of freedom, *avgyr* the average year of study period, *devg* equals 1 if data are for developing country, *trans* equals 1 if data are for transition country, *cs* equals 1 if data are cross-sectional, *growth* equals 1 if response variable is output growth, *industry* equals 1 if data are industry-level, *secdum* equals 1 if industry dummies are used, *empl* equals 1 if MNC presence is measured in employment, *assets* for foreign presence measured in assets and *output* for foreign presence measured in output.

Table 3.13: Panel meta-regression, all studies

	Random effects, including outliers				Random effects, excluding outliers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ldf	0.351	0.298			0.0828	0.0141		
	(1.29)	(1.17)			(0.60)	(0.12)		
avgyr	0.122*	0.0754			−0.00560	−0.0202		
	(2.02)	(1.49)			(−0.18)	(−0.60)		
devg	−1.752	−2.493			−0.247	−0.662		
	(−0.80)	(−1.50)			(−0.47)	(−1.06)		
trans	0.619	−0.243		2.345	0.727	−0.296		0.874
	(0.20)	(−0.13)		(1.12)	(0.89)	(−0.43)		(1.31)
cs	6.330**	5.452**		4.859*	1.993**	2.182**		1.819**
	(2.68)	(2.95)		(2.46)	(3.10)	(2.89)		(3.47)
growth	0.837		0.284	1.000	0.756		0.582	0.651
	(0.57)		(0.18)	(0.69)	(1.47)		(1.04)	(1.34)
industry	0.898		1.552	−0.747	1.787**		2.313**	1.467*
	(0.55)		(1.32)	(−0.41)	(2.74)		(4.02)	(2.44)
secdum	1.427		1.111		0.353		0.230	
	(1.16)		(0.69)		(0.54)		(0.37)	
empl	2.066		1.446	2.226	1.808*		1.680*	1.492**
	(0.58)		(0.51)	(1.03)	(2.42)		(2.42)	(2.61)
assets	−0.459		−0.255		0.577		0.537	
	(−0.16)		(−0.08)		(0.74)		(0.62)	
output	0.0349		0.596	0.378	1.505		1.829*	1.215†
	(0.01)		(0.23)	(0.22)	(1.60)		(2.13)	(1.77)
Constant	−246.2*	−151.5	0.189	−1.871	8.379	40.39	−1.426†	−1.691*
	(−2.04)	(−1.50)	(0.07)	(−1.21)	(0.13)	(0.60)	(−1.77)	(−2.39)
Observations	97	97	97	97	87	87	87	87
R^2	0.166	0.129	0.021	0.130	0.335	0.220	0.222	0.327

Notes: ldf is a logarithm of degrees of freedom, *avgyr* the average year of study period, *devg* equals 1 if data are for developing country, *trans* equals 1 if data are for transition country, *cs* equals 1 if data are cross-sectional, *growth* equals 1 if response variable is output growth, *industry* equals 1 if data are industry-level, *secdum* equals 1 if industry dummies are used, *empl* equals 1 if MNC presence is measured in employment, *assets* for foreign presence measured in assets and *output* for foreign presence measured in output.

Table 3.14: Probability meta-regression, all studies

	Probit—POSIT, all observations				Probit—SIGNIF, all observations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ldf	0.0666	0.00737			0.260**	0.144*		0.146*
	(0.79)	(0.11)			(2.78)	(2.13)		(2.15)
avgyr	−0.0195	−0.00244			−0.0640*	−0.0498*		−0.0561**
	(−0.54)	(−0.10)			(−2.51)	(−2.34)		(−2.62)
devg	0.264	−0.0107			0.427	0.261		
	(0.53)	(−0.03)			(0.95)	(0.74)		
trans	0.635	−0.290		0.397	0.413	0.153		
	(1.08)	(−0.74)		(1.01)	(0.74)	(0.40)		
cs	1.123*	1.100**		0.982*	0.873*	0.680†		0.630†
	(2.46)	(2.72)		(2.35)	(2.10)	(1.87)		(1.75)
growth	0.162		0.0380	0.202	0.378		0.317	
	(0.46)		(0.12)	(0.59)	(1.12)		(1.04)	
industry	1.602*		1.528**	1.311*	0.667		0.366	
	(2.27)		(2.77)	(2.13)	(1.30)		(0.99)	
secdum	0.297		0.0868		0.0283		−0.310	−0.0592
	(0.86)		(0.28)		(0.08)		(−1.03)	(−0.19)
empl	1.411*		1.114**	1.344**	0.0173		0.293	
	(2.45)		(2.63)	(2.60)	(0.03)		(0.76)	
assets	0.695		0.611	0.785†	0.197		0.423	0.485
	(1.42)		(1.41)	(1.65)	(0.44)		(1.03)	(1.53)
output	0.841		0.900*	0.959†	−0.738		0.0651	
	(1.62)		(1.97)	(1.93)	(−1.41)		(0.15)	
Constant	37.03	5.059	−0.558	−1.088*	124.6*	97.63*	−0.467	110.3**
	(0.51)	(0.10)	(−1.31)	(−1.97)	(2.47)	(2.31)	(−1.16)	(2.59)
Observations	97	97	97	97	97	97	97	97
Pseudo R^2	0.238	0.126	0.176	0.224	0.190	0.123	0.044	0.137

Notes: POSIT equals 1 if model reports positive spillovers and zero otherwise. *SIGNIF* equals 1 if model reports significant spillovers and zero otherwise. *ldf* is a logarithm of degrees of freedom, *avgyr* the average year of study period, *devg* equals 1 if data are for developing country, *trans* equals 1 if data are for transition country, *cs* equals 1 if data are cross-sectional, *growth* equals 1 if response variable is output growth, *industry* equals 1 if data are industry-level, *secdum* equals 1 if industry dummies are used, *empl* equals 1 if MNC presence is measured in employment, *assets* for foreign presence measured in assets and *output* for foreign presence measured in output.

Chapter 4

Demand for Gasoline Is More Price-Inelastic than Commonly Thought

Abstract

One of the most frequently examined statistical relationships in energy economics has been the price elasticity of gasoline demand. We conduct a quantitative survey of the estimates of elasticity reported for various countries around the world. Our meta-analysis indicates that the literature suffers from publication selection bias: insignificant or positive estimates of the price elasticity are rarely reported, although implausibly large negative estimates are reported regularly. In consequence, the average published estimates of both short- and long-run elasticities are exaggerated twofold. Using mixed-effects multilevel meta-regression, we show that after correction for publication bias the average long-run elasticity reaches -0.31 and the average short-run elasticity only -0.09.

Keywords: Gasoline demand, price elasticity, meta-analysis, publication selection bias

JEL Codes: C83, Q41, Q48

This paper is a joint work with Karel Janda and Tomas Havranek. We thank Martijn Brons and two anonymous reviewers for helpful comments on an earlier draft of the paper. We acknowledge financial support of the Grant Agency of Charles University (grant #76810), the Grant Agency of the Czech Republic (grant #P402/11/0948), and research project MSM0021620841. The paper has been published in Energy Economics.

4.1 Introduction

For the purposes of government policy concerning energy security, optimal taxation, and climate change, precise estimates of the price elasticity of gasoline demand are of principal importance. For example, if gasoline demand is highly price-inelastic, taxes will be ineffective in reducing gasoline consumption and the corresponding emissions of greenhouse gases. During the last 30 years the topic has attracted a lot of attention of economists who produced a plethora of empirical estimates of both short- and long-run price elasticities. Yet the estimates vary broadly.

A systematic method how to make use of all this work is to collect these numerous estimates and summarize them quantitatively. The method is called meta-analysis (Stanley 2001) and has long been used in economics following the seminal contribution by Stanley & Jarrell (1989). Recent applications of meta-analysis in economics include, among others, Card *et al.* (2010) on the evaluation of active labor market policy, Havranek (2010) on the trade effect of currency unions, and Horvathova (2010) on the impact of environmental performance on corporate financial performance.

Two international meta-analyses of the elasticity of gasoline demand have been conducted (Espey 1998; Brons *et al.* 2008). These meta-analyses study carefully the causes of heterogeneity observed in the literature. The average short- and long-run elasticities found by these meta-analyses were -0.26 and -0.58 (Espey 1998) and -0.34 and -0.84 (Brons *et al.* 2008). None of the meta-analyses, however, corrected the estimates for publication bias. It is well-known that publication selection can seriously bias the estimates of price elasticities because positive estimates are usually inconsistent with theory: for instance, Stanley (2005) documents how the price elasticity of water demand is exaggerated *fourfold* because of publication bias.

Publication selection bias, long recognized as a serious issue in empirical economics research (DeLong & Lang 1992; Card & Krueger 1995; Ashenfelter & Greenstone 2004), arises when statistically significant estimates or estimates with a particular sign are preferentially selected for publication. The bias stems from the preference of authors, editors, or reviewers for results that tell a story and are theory-consistent.

Publication bias has been found in many areas of empirical economics (Doucouliagos & Stanley 2008).

The effects of publication selection differ at the study and literature levels. At the study level it is reasonable not to base discussion on the estimates of the price elasticity of gasoline demand that are positive—few would consider gasoline to be a Giffen good, and positive estimates are thus most likely due to misspecifications. On the other hand, it is far more difficult to identify large negative estimates that are also due to misspecifications. If all researchers discard positive estimates of the price elasticity but keep large negative estimates, the average impression derived from the literature will be biased toward stronger elasticity. Thus, at the literature level the mean estimate must be corrected for publication bias.

We employ recently developed meta-analysis methods to test for publication bias and estimate the corrected elasticity beyond. The mixed-effects multilevel meta-regression takes into account heteroscedasticity, which is inevitable in meta-analysis, and between-study heterogeneity, which is likely to occur in most areas of empirical economics. We do not, however, investigate heterogeneity explicitly, as this issue was thoroughly examined by the two previous meta-analyses.

The paper is structured as follows. Section 4.2 discusses the process of selecting studies to be included in the meta-analysis and the properties of the data. Section 4.3 describes the meta-analysis methods used to detect and correct for publication bias. Section 4.4 discusses the results of the meta-regression. Section 4.5 concludes.

4.2 The Elasticity Estimates Data Set

The first step of meta-analysis is the collection of primary studies. We examined all studies used by the most recent meta-analysis (Brons *et al.* 2008), but because the sample used by Brons *et al.* (2008) ends in 1999, we additionally searched the EconLit and Scopus databases for new studies published between 2000 and 2011. To be able to use modern meta-analysis methods and correct for publication bias, we need the standard error of each estimate of elasticity; therefore we have to exclude studies that do not report standard errors (or any other statistics from which standard errors could be computed). Concerning the definition of short- and long-term elasticity estimates, we follow the approach described in the first meta-analysis on this topic, Espey (1998).

Some meta-analysts argue for using estimates from all available studies in hope that the inclusion of unpublished studies will alleviate publication bias. Nevertheless, rational authors of primary studies are likely to polish even early drafts of their papers as they prepare for journal submission, or may use the intuitive sign of the estimate as a specification check. in a large survey of economics meta-analyses, Doucouliagos & Stanley (2008) document that the inclusion of working papers does not help mitigate publication bias. Hence we follow, among others, Abreu *et al.* (2005) and collect estimates only from studies published in peer-reviewed journals—as a simple criterion of quality.[1] In sum, our sample consists of 202 estimates of the price elasticity of gasoline demand taken from 41 journal articles.

Table 4.1: List of primary studies used

Abdel-Khalek (1988)	Drollas (1984)	Pock (2010)
Akinboade *et al.* (2008)	Eltony (1993)	Ramanathan (1999)
Alves & Bueno (2003)	Eltony & Al-Mutairi (1995)	Ramsey *et al.* (1975)
Archibald & Gillingham (1980)	Gallini (1983)	Reza & Spiro (1979)
Archibald & Gillingham (1981)	Houthakker *et al.* (1974)	Sipes & Mendelsohn (2001)
Baltagi & Griffin (1983)	Iwayemi *et al.* (2010)	Sterner (1991)
Baltagi & Griffin (1997)	Kennedy (1974)	Storchmann (2005)
Bentzen (1994)	Kim *et al.* (2011)	Tishler (1983)
Berndt & Botero (1985)	Kraft & Rodekohr (1978)	Uri & Hassanein (1985)
Berzeg (1982)	Kwast (1980)	Wadud *et al.* (2009)
Crôtte *et al.* (2010)	Lin *et al.* (1985)	West & Williams III (2007)
Dahl (1978)	Manzan & Zerom (2010)	Wheaton (1982)
Dahl (1979)	Mehta *et al.* (1978)	Wirl (1991)
Dahl (1982)	Nicol (2003)	

All studies included in our meta-analysis are listed in Table 4.1. The oldest study in our sample was published in 1974 and the most recent in 2011. Energy Economics appears to be the primary outlet for this literature—13 studies, one third

[1] It should be noted, however, that some meta-analyses find a significant difference in the magnitude of publication bias between published and unpublished studies (for example, Havranek & Irsova 2011). Fortunately, provided with a sufficient number of estimates, modern meta-analysis methods allow us to filter out publication bias regardless of its magnitude.

of the entire usable literature, were published in Energy Economics, as well as both previous meta-analyses of the elasticity of gasoline demand.

Out of the 202 estimates we collected, 110 are short-run elasticities and 92 long-run ones. Summary statistics for these estimates of elasticities are reported in Table 4.2: the estimates of the short-run elasticity range from -0.96 to 0.08 with the mean estimate -0.23; the estimates of long-run elasticity range from -1.59 to -0.10 with the mean estimate reaching -0.69. Thus the simple averages of the estimates of both the short- and long-run elasticity in our sample are close to those reported by the earlier meta-analyses (Espey 1998; Brons *et al.* 2008). If there is publication selection bias against positive (or insignificant negative) estimates of elasticities, however, these simple averages will overstate the true elasticity.

Table 4.2: Summary statistics

Variable	Observations	Mean	Median	Std. dev.	Min	Max
Short-run elasticity	110	-0.227	-0.190	0.158	-0.960	0.080
Long-run elasticity	92	-0.691	-0.632	0.332	-1.590	-0.102

Figure 4.1: Kernel density of the estimated elasticities

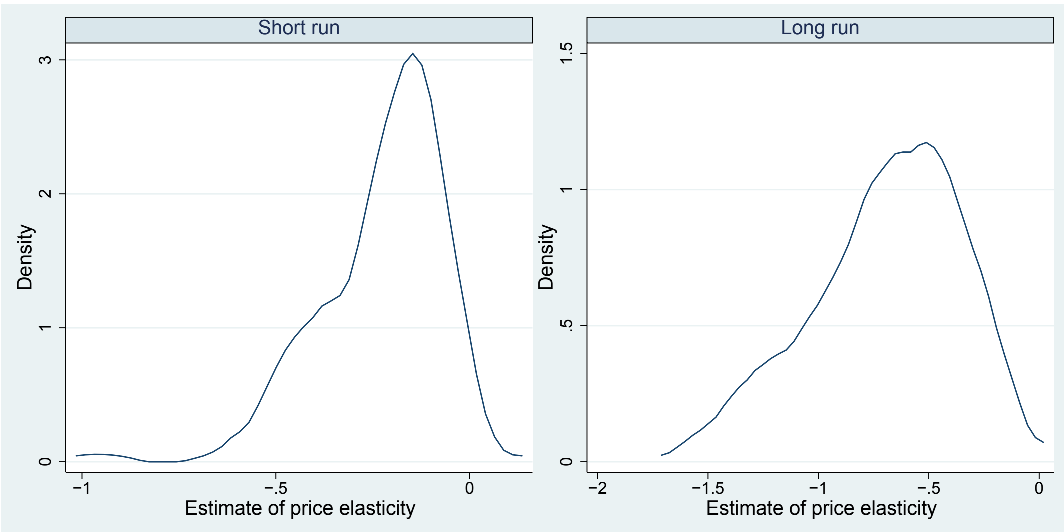


Figure 4.1 depicts the kernel density of the estimates of short- and long-run elasticities; we use the Epanechnikov kernel in the estimation. It is apparent that

both distributions are strongly skewed. Positive estimates of the price elasticity of gasoline demand are rarely published, so that the negative (that is, left-hand-side) tails of the distributions get much heavier. This suggests that something more than pure sampling error is driving the distribution of the results: by no means are they distributed normally around a hypothetical true effect, which is also confirmed by goodness-of-fit tests. Normal distribution of the estimated elasticities in the absence of publication bias is a standard assumption in meta-analysis (Stanley 2005; 2008), which stems from the fact that individual researchers estimate elasticities as regression parameters (assuming t-distribution, which is close to normal in large samples). Nevertheless, more specialized methods are needed to establish evidence for the presence of publication bias.

4.3 Meta-Analysis Methodology

A common method of assessing publication bias is an examination of the so-called funnel plot (Sutton *et al.* 2000; Stanley & Doucouliagos 2010). The funnel plot depicts the estimated elasticity on the horizontal axis against the precision of the estimate of elasticity (the inverse of the standard error) on the vertical axis. The most precise estimates will be close to the true effect, but the less precise ones will be more dispersed; in consequence the cloud of estimates should resemble an inverted funnel. When the literature is free of publication bias the funnel will be symmetrical around the values with the highest precision since all imprecise estimates of elasticity will have the same chance of being reported. While the funnel plot is a useful device, formal econometric methods are needed to estimate precisely the true elasticity beyond publication bias.

In the absence of publication bias the estimates of elasticities are randomly distributed around the true mean elasticity, e_0. Nevertheless, if some estimates end in the "file drawer" (Rosenthal 1979) because they are insignificant or have a positive sign, the reported estimates will be correlated with their standard errors (Card &

Krueger 1995; Ashenfelter *et al.* 1999):

$$e_i = e_0 + \beta_0 \cdot Se(e_i) + u_i, \quad u_i|Se(e_i) \sim N(0, \delta^2), \tag{4.1}$$

where e_i denotes the estimate of elasticity, e_0 is the average underlying elasticity, $Se(e_i)$ is the standard error of e_i, β_0 measures the magnitude of publication bias, and u_i is a disturbance term. For example, if a statistically significant effect is required, an author who has few observations may run a specification search until the estimate becomes large enough to offset the high standard errors. Specification (4.1) can also be interpreted as a test of the asymmetry of the funnel plot; it follows from rotating the axes of the plot and inverting the values on the new horizontal axis. A significant estimate of β_0 then provides formal evidence for funnel asymmetry. Because specification (4.1) is likely heteroscedastic (the explanatory variable is a sample estimate of the standard deviation of the response variable), in practice it is usually estimated by weighted least squares (Stanley 2005; 2008):

$$e_i/Se(e_i) = t_i = e_0 \cdot 1/Se(e_i) + \beta_0 + \xi_i, \quad \xi_i|Se(e_i) \sim N(0, \sigma^2). \tag{4.2}$$

Monte Carlo simulations and many recent meta-analyses suggest that this parsimonious specification is also effective in testing the significance of the true elasticity beyond publication bias, coefficient e_0 (Stanley 2008).

In meta-analysis we have to take into consideration that estimates coming from one study are likely to be dependent. A common way how to cope with this problem is to employ the mixed-effects multilevel model (Doucouliagos & Stanley 2009), which allows for unobserved between-study heterogeneity. Between-study heterogeneity is likely to be substantial since in our case the primary studies use data from different countries. We specify the model following Havranek & Irsova (2011):

$$t_{ij} = e_0 \cdot 1/Se(e_{ij}) + \beta_0 + \zeta_j + \epsilon_{ij}, \quad \zeta_j|Se(e_{ij}) \sim N(0, \psi), \quad \epsilon_{ij}|Se(e_{ij}), \zeta_j \sim N(0, \theta), \tag{4.3}$$

where i and j denote estimate and study subscripts. The overall error term (ξ_{ij}) now breaks down into study-level random effects (ζ_j) and estimate-level disturbances (ϵ_{ij}). The variance of these error terms is additive because both components are assumed to be independent: $\text{Var}(\xi_{ij}) = \psi + \theta$, where ψ denotes between-study variance (that is, between-study heterogeneity) and θ within-study variance. When ψ approaches zero the benefit of using the mixed-effect multilevel estimator instead of simple ordinary least squares (OLS) becomes negligible; we will use likelihood-ratio tests to examine this condition.

The mixed-effects multilevel model is analogous to the random-effects model commonly used in panel-data econometrics. The terminology, however, follows hierarchical data modeling: the model is called "mixed-effects" since it contains a fixed (e_0) as well as a random part (ζ_j). For the purposes of meta-analysis the multilevel framework is more suitable because it takes into account the unbalancedness of the data (the maximum likelihood estimator is used instead of generalized least squares) and allows for nesting multiple random effects (author-, study-, or country-level), and is thus more flexible.

The high degree of unbalancedness of the data in meta-analysis makes a reliable testing of the exogeneity assumptions behind the mixed-effects model difficult; fixed effects in the panel-data sense are generally inappropriate for meta-analysis since some studies report only one usable estimate. We follow the recommendation of an authoritative survey of meta-analyses in environmental and resource economics (Nelson & Kennedy 2009, p. 358): "The advantages of random-effects estimation [in meta-analysis] are so strong that this estimation procedure should be employed unless a very strong case can be made for its inappropriateness." As a robustness check, however, we also employ OLS with clustered standard errors. Large differences between the estimates based on OLS and on mixed effects may signal a violation of the exogeneity assumptions.

Specification (4.3) enables us to examine the significance and magnitude of publication bias (β_0) and the *significance* of the true elasticity beyond publication bias (e_0). To examine the *magnitude* of the true elasticity, Stanley & Doucouliagos (2007; 2011) recommend an augmented version of (4.3); this specification is also supported as the best method to correct for publication bias by a survey of meta-analysis methods published in the British Medical Journal (Moreno *et al.* 2009). The specification is based on the assumption that the relation between standard errors and publication bias in (4.1) is quadratic; the model is called the Heckman meta-regression (see Stanley & Doucouliagos 2007, for details). When heteroscedasticity and between-study heterogeneity are taken into account, the specification assumes the following form:

$$t_{ij} = e_0 \cdot 1/Se(e_{ij}) + \beta_0 SE + \zeta_j + \epsilon_{ij}, \quad \zeta_j | Se(e_{ij}) \sim N(0, \psi), \quad \epsilon_{ij} | Se(e_{ij}), \zeta_j \sim N(0, \theta), \tag{4.4}$$

where e_0 measures the magnitude of the average elasticity corrected for publication bias.

In this paper we concentrate on the average estimate of elasticity and do not investigate the sources of heterogeneity in the estimates since heterogeneity was carefully examined by the previous meta-analyses. Also the measure of publication selection bias estimated in specification (4.3) is mean across all countries and methods used for estimation in primary studies. Nevertheless, it would be useful to find out whether some aspects of primary studies are associated with more publication bias than others. For this exercise we select three aspects identified as important for the differences in reported estimates by the previous meta-analyses: the use of US against non-US data, the use of time-series against cross-sectional data, and study publication date. We employ the methodology of Stanley *et al.* (2008), who interact publication bias and study aspects in meta-regression (4.1). After weighting by the standard error and adding study-level random effects the specification becomes

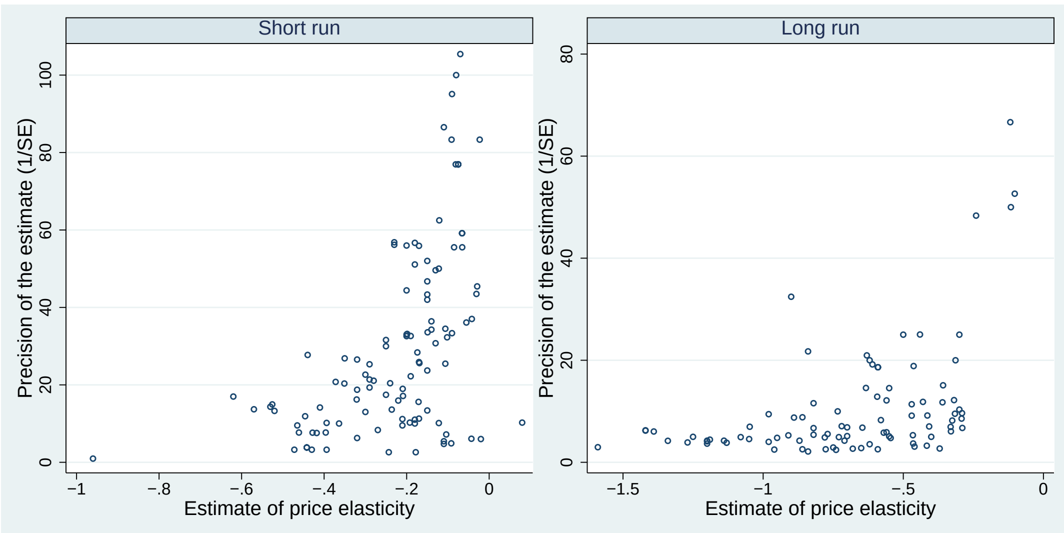
$$t_{ij} = e_0 \cdot 1/Se(e_{ij}) + \alpha_1 usdata_{ij} + \alpha_2 csection_{ij} + \alpha_3 pubdate_j + \beta_0 + \zeta_j + \epsilon_{ij}, \tag{4.5}$$

where *usdata* is a dummy variable that equals one if the primary study uses data for the US to estimate the particular elasticity and zero otherwise, *csection* is a dummy variable that equals one if the primary study uses data with a cross-sectional dimension (including panel data) and zero otherwise, *pubdate* denotes the year of publication of the primary study, and other variables have the same properties as in specification (4.3).

4.4 Results

Figure 4.2 depicts funnel plots for the estimates of short- and long-run price elasticities of gasoline demand. The funnels are heavily asymmetrical: the right-hand part of the funnels is almost completely missing, hence we have a good reason to believe that publication selection bias in this literature is strong. The estimates with the highest precision are negative but small in magnitude, positive estimates are almost never published, while imprecise negative estimates are published regularly—therefore the average reported estimate is likely to be biased downwards.

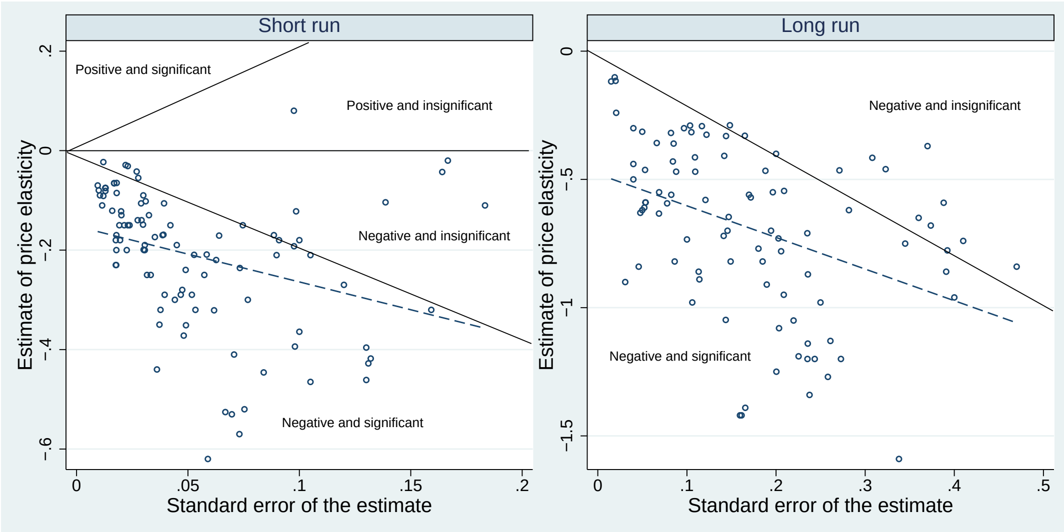
Figure 4.2: Funnel plot of the estimated elasticities



Notes: In the absence of publication bias the funnel should be symmetrical around the most precise estimates of elasticities. This funnel is asymmetrical, which suggests that positive estimates are not reported, even though we should observe a few of them in the literature due to the laws of chance.

The formal test of publication bias, described by regression (4.1), follows directly from the funnel plot—hence, it is often called the funnel asymmetry test. We illustrate the transition from the funnel plot to the funnel asymmetry test in Figure 4.3. In this scatter plot the size of the estimates of elasticities is depicted on the vertical

Figure 4.3: Visualization of the funnel asymmetry test



Notes: The solid lines denote the combinations of the estimates of elasticities and their standard errors for which the t-statistic in the absolute value equals two. The dashed line denotes a linear fit of the points [that is, regression (4.1) or, in other words, the funnel asymmetry test]; its negative slope suggests publication bias.

axis; the horizontal axis measures the standard errors of the estimates. (Compared with the funnel plot, the axes are switched and the values on the new horizontal axis are inverted.) For the short-run elasticities, a few estimates with extremely large standard errors are cut from the figure so that the overall pattern can be seen. Now, if we interpret Figure 4.3 as a regression relationship, we get equation (4.1).

Nevertheless, in regression (4.1) publication bias is only related to the standard error and, thus, seemingly only to statistical significance. It remains to be shown that this test captures both sources of publication bias: the one stemming from the selection of the significant estimates (type II bias in the terminology of Stanley 2005) and the one stemming from the selection of the estimates with an intuitive sign (type I bias). The suitability of the funnel asymmetry test to filter out both sources of publication bias is often stated in the meta-analysis literature (for instance, Doucouliagos & Paldam 2009), but rarely discussed in detail.[2]

If no publication bias was present, the observations in Figure 4.3 would form an isosceles triangle with the tip pointing to the most precise estimate of the elasticity;

[2]We thank an anonymous reviewer for pointing out this problem.

regressing the estimates on their standard errors would yield no statistically significant slope coefficient. First, let us suppose that only negative estimates, irrespective of their statistical significance, were reported. In such a case the triangle would lose its upper part. Regression (4.1) would yield a negative slope coefficient, evidence of publication bias. Second, let us assume that researchers suppress estimates insignificant at the 5% level, irrespective of the sign. In Figure 4.3 we depict the boundary of significance at the 5% level: the solid lines show the combinations of the magnitudes of estimates and their standard errors for which the corresponding t-statistic equals two in the absolute value. In the case of type II publication bias the triangle would lose its middle part (no estimates with $|t| < 2$ would be reported). Because the true elasticity is most likely negative, few positively significant estimates would appear, and regression (4.1) would again yield a negative slope coefficient, a sign of publication bias.[3]

Figure 4.3, especially the left panel depicting short-run elasticities, suggests that both sources of publication bias play a role in the empirical literature on gasoline demand. Statistical significance is important; insignificant negative estimates of price elasticities seem to be much less likely to get published than the significant negative estimates. In the absence of type II publication bias, the negative estimates should be approximately symmetrical with respect to the $t = -2$ line, but in this case the insignificant estimates are apparently underrepresented.

The sign of the estimate is important as well. In the absence of type I publication bias, all estimates should be symmetrical with respect to the $e = e_0$ line, where e_0 denotes the true elasticity (approximately -0.09 for the short run, as will be showed later). In such a case, we should observe more positive estimates, including a few significant ones due to the laws of chance. There are, however, unlikely to be any positive but precise estimates of price elasticities. In fact, if the true elasticity was large enough and the estimation methods were precise enough, we would only observe negative estimates of price elasticities even in the absence of publication bias. But

[3]If the true elasticity was zero and there was no preference for negative estimates, the funnel asymmetry test could not detect type II publication bias. But then the selection would be symmetrical and would not bias the arithmetic average of elasticities taken from the literature.

then the funnel plot would be symmetrical, although all observed estimates would have the same sign, as noted by Doucouliagos & Stanley (2009); this is not the case of Figure 4.2.

The dotted line in Figure 4.3 shows a linear fit based on regression (4.1): its negative slope indicates publication bias. In the end the regression computes the average estimate of elasticity conditional on the standard error being close to zero. In other words, it looks for a hypothetical estimate with infinite precision, and in Figure 4.3 the infinitely precise estimate would be represented by the intercept of the dotted line with the vertical axis. Nevertheless, as discussed in Section 4.3, because of heteroscedasticity and between-study heterogeneity regression (4.1) is rarely estimated itself.

Table 4.3 summarizes the results of a regression based on specification (4.3) [the mixed-effects weighted-least-squares version of (4.1)]. The regression is estimated separately for the short- and long-run elasticity to obtain precise estimates of these individual elasticities in the later stage of our analysis. Likelihood-ratio tests reject the null hypothesis, which suggests that between-study heterogeneity is substantial, the OLS is misspecified, and the mixed-effects model is thus more reliable. Moreover the differences between the OLS and the mixed-effects model are small, indicating that the exogeneity assumptions behind the mixed-effects model are not seriously violated. We also estimated several nested models with additional author- and country-level random effects, but according to likelihood-ratio tests these models do not significantly differ from the baseline model that only accounts for between-study heterogeneity.

As expected after examining the funnel plots, the meta-regression identifies downward publication bias, significant at the 1% level for all specifications. In all cases the intensity of publication bias, β_0, is also larger than two in the absolute value. According to Doucouliagos & Stanley (2008), such magnitude of publication bias is considered "severe" and signals serious selection efforts: if the true elasticity was zero and only significantly negative estimates were reported, the estimated coefficient for publication bias would approach two, the most commonly used critical value of the

t-statistic. Publication bias in this literature is hence strong enough to produce a significant average estimate of the effect even if there was none in reality.

Table 4.3: Test of publication bias

	Mixed-effects multilevel		Clustered OLS	
Response variable: t-statistic	Short run	Long run	Short run	Long run
Constant (publication bias)	-2.587***	-2.491***	-2.890***	-3.570***
	(0.465)	(0.707)	(0.595)	(0.808)
$1/SE$	-0.0611***	-0.237***	-0.0651***	-0.189*
	(0.0111)	(0.0393)	(0.0152)	(0.111)
Observations	110	92	110	92
Likelihood-ratio test (χ^2)	21.78***	19.71***		

Notes: Standard errors, clustered at the study level for OLS, in parentheses. Null hypothesis for the likelihood-ratio test: no between-study heterogeneity (that is, the mixed-effects multilevel model has no benefit over OLS). ***, **, and * denote significance at the 1%, 5%, and 10% levels.

Nevertheless, Table 4.3 also shows that the estimate of the true effect (the coefficient for $1/SE$) is significant at least at the 10% level for all specifications; it is significant even at the 1% level in our preferred mixed-effects model. Thus, on average, both the short- and long-run price elasticity of gasoline demand is statistically different from zero even after correcting for publication bias. To estimate the true average elasticity precisely, we need to employ the Heckman meta-regression proposed by Stanley & Doucouliagos (2007) and corroborated by Moreno *et al.* (2009). This is achieved by estimating regression (4.4); the results are reported in Table 4.4. Similarly to the previous case, likelihood-ratio tests suggest that the OLS is misspecified, and we therefore only discuss the results of the mixed-effects model.

After correcting for publication bias, our best estimate indicates that the mean short-run elasticity reaches -0.09 with a 95% confidence interval (-0.12, -0.07). The corrected estimate of the long-run elasticity reaches -0.31 with a 95% confidence interval (-0.38, -0.25). This sharply contrasts to the simple uncorrected averages amounting to -0.23 and -0.69: publication bias exaggerates the average reported elasticity more than twofold. For instance, concerning the short-run elasticity, only 18 out of the 110 estimates we collected are smaller in the absolute value than the true average effect (-0.09). Therefore as much as 74 positive (or negative but insignificant) estimates of the short-run price elasticitiy of gasoline were likely not reported because of publication selection. In other words, about 40% of all estimated elasticities may be put into the "file drawer."

Table 4.4: Test of the true elasticity beyond publication bias

	Mixed-effects multilevel		Clustered OLS	
Response variable: t-statistic	Short run	Long run	Short run	Long run
$1/SE$ (true elasticity)	-0.0913***	-0.314***	-0.120***	-0.307**
	(0.0120)	(0.0334)	(0.0145)	(0.115)
SE	-0.975	-2.396	-4.960*	-9.343***
	(2.094)	(2.668)	(2.558)	(3.054)
Observations	110	92	110	92
Likelihood-ratio test (χ^2)	37.28***	34.45***		

Notes: Standard errors, clustered at the study level for OLS, in parentheses. Null hypothesis for the likelihood-ratio test: no between-study heterogeneity (that is, the mixed-effects multilevel model has no benefit over OLS). ***, **, and * denote significance at the 1%, 5%, and 10% levels.

Since our preferred mixed-effects estimator not only corrects for publication bias, but also involves several other adjustments (for example, weighted least squares specification or study-specific random effects), the comparison with a simple average may not be straightforward, however. As a robustness check, using the mixed-effects model we also estimate average elasticities not corrected for publication bias. The uncorrected averages reported by mixed effects are -0.23 and -0.63 for short- and long-run elasticities; that is, very close to the simple averages (-0.23 and -0.69). Given these results, we argue that the difference between corrected estimates and simple averages is due to publication bias and not specification characteristics of the meta-regression.[4]

To test whether the degree of publication selection depends on study aspects, we estimate specification (4.5) and report the results in Table 4.5. In this case publication bias is no more represented by the constant only; the bias is captured by all variables with the exception of $1/SE$. The coefficient for $1/SE$ still represents the true effect corrected for publication bias, and the results suggest that corrected estimates in Table 4.5 are very similar to those in Table 4.3 even though we now

[4]We thank Martijn Brons for pointing out this problem.

Table 4.5: Multivariate meta-regression

	Mixed-effects multilevel		Clustered OLS	
Response variable: t-statistic	Short run	Long run	Short run	Long run
$1/SE$	-0.0547***	-0.228***	-0.0709***	-0.231***
	(0.0124)	(0.0318)	(0.0155)	(0.0770)
US data	0.375	1.964**	1.090	2.697**
	(0.756)	(0.942)	(0.654)	(1.089)
Cross-sectional dimension	-1.270*	-2.142***	0.170	-1.958***
	(0.769)	(0.795)	(1.217)	(0.645)
Year of publication	0.0130	0.0796**	0.0366	0.104**
	(0.0370)	(0.0393)	(0.0487)	(0.0475)
Constant	-27.99	-160.7**	-75.95	-210.5**
	(73.80)	(78.42)	(97.30)	(94.37)
Observations	110	92	110	92
Test of joint significance	3.47	18.26***	0.94	3.70**

Notes: Standard errors, clustered at the study level for OLS, in parentheses. Null hypothesis for the test of joint significance: $\alpha_1 = \alpha_2 = \alpha_3 = 0$ [see regression (4.5)]; Wald test is used for the mixed-effects model, F-test for OLS. ***, **, and * denote significance at the 1%, 5%, and 10% levels.

control for several study aspects. For the short-run estimates of elasticity the test of joint significance does not reject the hypothesis that the pattern of publication is the same for various study aspects. For the long-run estimates, however, the differences are statistically significant. The use of US data is associated with less publication bias,[5] while the use of data with a cross-sectional dimension is associated with more bias. Finally, the magnitude of publication bias decreases in time, which is consistent with the economics-research-cycle hypothesis (Goldfarb 1995; Stanley *et al.* 2008).

4.5 Conclusion

We conduct a quantitative survey of journal articles estimating the price elasticity of gasoline demand. In contrast to previous meta-analyses on this topic, we take into account publication selection bias using the mixed-effects multilevel meta-regression. Publication bias in this area is strong; when we correct for the bias, we obtain estimates of short- and long-run elasticities that are approximately *half*, compared to

[5] Publication selection creates a downward bias among the estimates of price elasticities, so a positive estimated coefficient on the interaction between the use of US data and standard error means less downward bias when US data we used.

the results of the previously published meta-analyses and also to the simple mean of all estimates in our sample of literature. If the simple mean reflects our profession's impression about the magnitude of the price elasticity of gasoline demand, the impression exaggerates the true elasticity twofold.

This paper complements the previously published meta-analyses on the price elasticity of gasoline demand (Espey 1998; Brons *et al.* 2008). These meta-analyses focus on the reasons why estimates of elasticities differ for different regions and different methods used and provide mean estimates of short- and long-run price elasticities as a bonus. It is important to bear in mind the differences between the methods used in this paper to deliver the average estimates of elasticity and the methods used in Espey (1998) and Brons *et al.* (2008). First, the estimates of Brons *et al.* (2008) are based on a seemingly unrelated regression model with cross-equation restrictions. Second, neither Espey (1998) nor Brons *et al.* (2008) use a multilevel approach to distinguish between study-level and estimate-level variation. Third, the sets of studies differ among the three meta-analyses. Although the estimates of average elasticity are therefore not directly comparable, we argue there is a strong case for the presence of publication bias in favor of larger negative estimates of elasticities in the literature.

The estimated elasticities corrected for publication bias, -0.09 for the short run and -0.31 for the long run, are average across many countries, methods, and time periods; we report them as reference values. A similar pattern of publication bias, however, is likely to appear in any subset of the literature. Thus large negative estimates of price elasticities should be taken with a grain of salt.

Concerning future research, authors interested in figures for individual countries may collect more estimates from working papers, dissertations, and other mimeographs, which should provide enough degrees of freedom to estimate the price elasticity of gasoline demand for each country using the methodology described in this paper. Next, since previous meta-analyses suggest that study design may affect results in a systematic way, researchers could define best-practice methodology and estimate price elasticities conditional on such best practice to filter out the effects of misspecifications. Finally, given the number of studies conducted on this topic each year, in the meta-analysis framework it is also possible to test whether the price elasticity of gasoline demand changed during the last decade when the prices of petroleum products surged.

References

Abdel-Khalek, G. (1988): "Income and price elasticities of energy consumption in Egypt: A time-series analysis." *Energy Economics* **10(1)**: pp. 47–58.

Abreu, M., H. L. F. de Groot, & R. J. G. M. Florax (2005): "A Meta-Analysis of β-Convergence: the Legendary 2%." *Journal of Economic Surveys* **19(3)**: pp. 389–420.

Akinboade, O. A., E. Ziramba, & W. L. Kumo (2008): "The demand for gasoline in South Africa: An empirical analysis using co-integration techniques." *Energy Economics* **30(6)**: pp. 3222–3229.

Alves, D. C. O. & R. D. L. d. S. Bueno (2003): "Short-run, long-run and cross elasticities of gasoline demand in Brazil." *Energy Economics* **25(2)**: pp. 191–199.

Archibald, R. & R. Gillingham (1980): "An Analysis of the Short-Run Consumer Demand for Gasoline Using Household Survey Data." *The Review of Economics and Statistics* **62(4)**: pp. 622–28.

Archibald, R. & R. Gillingham (1981): "The Distributional Impact of Alternative Gasoline Conservation Policies." *Bell Journal of Economics* **12(2)**: pp. 426–444.

Ashenfelter, O. & M. Greenstone (2004): "Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias." *American Economic Review* **94(2)**: pp. 454–460.

Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A review of estimates of the schooling/earnings relationship, with tests for publication bias." *Labour Economics* **6(4)**: pp. 453–470.

Baltagi, B. H. & J. M. Griffin (1983): "Gasoline demand in the OECD: An application of pooling and testing procedures." *European Economic Review* **22(2)**: pp. 117–137.

Baltagi, B. H. & J. M. Griffin (1997): "Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline." *Journal of Econometrics* **77(2)**: pp. 303–327.

Bentzen, J. (1994): "An empirical analysis of gasoline demand in Denmark using cointegration techniques." *Energy Economics* **16(2)**: pp. 139–143.

Berndt, E. R. & G. Botero (1985): "Energy demand in the transportation sector of Mexico." *Journal of Development Economics* **17(3)**: pp. 219–238.

Berzeg, K. (1982): "Demand for motor gasoline: a generalized error components model." *Southern Economic Journal* **49**: pp. 359–373.

Brons, M., P. Nijkamp, E. Pels, & P. Rietveld (2008): "A meta-analysis of the price elasticity of gasoline demand. A SUR approach." *Energy Economics* **30(5)**: pp. 2105–2122.

Card, D., J. Kluve, & A. Weber (2010): "Active labour market policy evaluations: A meta-analysis." *Economic Journal* **120(548)**: pp. F452–F477.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review* **85(2)**: pp. 238–43.

Crôtte, A., R. B. Noland, & D. J. Graham (2010): "An analysis of gasoline demand elasticities at the national and local levels in Mexico." *Energy Policy* **38(8)**: pp. 4445–4456.

Dahl, C. A. (1978): "American energy consumption—Extravagant or economical? A study of gasoline demand." *Resources and Energy* **1(4)**: pp. 359–373.

Dahl, C. A. (1979): "Consumer Adjustment to a Gasoline Tax." *The Review of Economics and Statistics* **61(3)**: pp. 427–32.

Dahl, C. A. (1982): "Do Gasoline Demand Elasticities Vary?" *Land Economics* **58(3)**: pp. 373–382.

DeLong, J. B. & K. Lang (1992): "Are All Economic Hypotheses False?" *Journal* *of Political Economy* **100(6)**: pp. 1257–72.

Doucouliagos, H. & M. Paldam (2009): "The Aid Effectiveness Literature: The Sad Results of 40 Years of Research." *Journal of Economic Surveys* **23(3)**: pp. 433–461.

Doucouliagos, H. & T. Stanley (2008): "Theory Competition and Selectivity: Are All Economic Facts Greatly Exaggerated?" *Economics Series Working Paper 06*, Deakin University.

Doucouliagos, H. & T. D. Stanley (2009): "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* **47(2)**: pp. 406–428.

Drollas, L. P. (1984): "The demand for gasoline: Further evidence." *Energy Economics* **6(1)**: pp. 71–82.

Eltony, M. (1993): "Transport gasoline demand in Canada." *Journal of Transport Economics and Policy* **27**: pp. 193–208.

Eltony, M. N. & N. H. Al-Mutairi (1995): "Demand for gasoline in Kuwait: An empirical analysis using cointegration techniques." *Energy Economics* **17(3)**: pp. 249–253.

Espey, M. (1998): "Gasoline demand revisited: an international meta-analysis of elasticities." *Energy Economics* **20(3)**: pp. 273–295.

Gallini, N. T. (1983): "Demand for Gasoline in Canada." *Canadian Journal of Economics* **16(2)**: pp. 299–324.

Goldfarb, R. S. (1995): "The Economist-as-Audience Needs a Methodology of Plausible Inference." *Journal of Economic Methodology* **2(2)**: pp. 201–22.

Havranek, T. (2010): "Rose Effect and the Euro: Is the Magic Gone?" *Review of World Economics* **146(2)**: pp. 241–261.

Havranek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85(2)**: pp. 234–244.

Horvathova, E. (2010): "Does environmental performance affect financial performance? A meta-analysis." *Ecological Economics* **70(1)**: pp. 52–59.

Houthakker, H., P. Verleger, & D. Sheehan (1974): "Dynamic demand analysis for gasoline and residential electricity." *American Journal of Agricultural Economics* **56**: p. 412–418.

Iwayemi, A., A. Adenikinju, & M. A. Babatunde (2010): "Estimating petroleum products demand elasticities in Nigeria: A multivariate cointegration approach." *Energy Economics* **32(1)**: pp. 73–85.

Kennedy, M. (1974): "An Economic Model of the World Oil Market." *Bell Journal of Economics* **5(2)**: pp. 540–577.

Kim, Y.-D., H.-O. Han, & Y.-S. Moon (2011): "The empirical effects of a gasoline tax on CO2 emissions reductions from transportation sector in Korea." *Energy Policy* **39(2)**: pp. 981–989.

Kraft, J. & M. Rodekohr (1978): "Regional demand for gasoline: A temporal cross-section specification." *Journal of Regional Science* **18**: pp. 45–56.

Kwast, M. (1980): "A note on the structural stability of gasoline demand and the welfare economics of gasoline taxation." *Southern Economic Journal* **46**: pp. 1212–1220.

Lin, A.-l., E. N. Botsas, & S. A. Monroe (1985): "State gasoline consumption in the USA: An econometric analysis." *Energy Economics* **7(1)**: pp. 29–36.

Manzan, S. & D. Zerom (2010): "A Semiparametric Analysis of Gasoline Demand in the United States Reexamining The Impact of Price." *Econometric Reviews* **29(4)**: pp. 439–468.

Mehta, J. S., G. V. L. Narasimham, & P. A. V. B. Swamy (1978): "Estimation of a dynamic demand function for gasoline with different schemes of parameter variation." *Journal of Econometrics* **7(3)**: pp. 263–279.

Moreno, S. G., A. J. Sutton, E. H. Turner, K. R. Abrams, N. J. Cooper, T. P. Palmer, & A. Ades (2009): "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**: pp. 494–498.

Nelson, J. & P. Kennedy (2009): "The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment." *Environmental & Resource Economics* **42(3)**: pp. 345–377.

Nicol, C. J. (2003): "Elasticities of demand for gasoline in Canada and the United States." *Energy Economics* **25(2)**: pp. 201–214.

Pock, M. (2010): "Gasoline demand in Europe: New insights." *Energy Economics* **32(1)**: pp. 54–62.

Ramanathan, R. (1999): "Short- and long-run elasticities of gasoline demand in India: An empirical analysis using cointegration techniques." *Energy Economics* **21(4)**: pp. 321–330.

Ramsey, J. B., R. Rasche, & B. T. Allen (1975): "An Analysis of the Private and Commercial Demand for Gasoline." *The Review of Economics and Statistics* **57(4)**: pp. 502–07.

Reza, A. & M. Spiro (1979): "The demand for passenger car transport services and for gasoline." *Journal of Transport Economics and Policy* **13**: pp. 304–319.

Rosenthal, R. (1979): "The 'file drawer problem' and tolerance for null results." *Psychological Bulletin* **86**: pp. 638–41.

Sipes, K. N. & R. Mendelsohn (2001): "The effectiveness of gasoline taxation to manage air pollution." *Ecological Economics* **36(2)**: pp. 299–309.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

STANLEY, T. D. & H. DOUCOULIAGOS (2007): "Identifying and Correcting Publication Selection Bias in the Efficiency-Wage Literature: Heckman Meta-Regression." *Economics Series 2007/11*, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

STANLEY, T. D. & H. DOUCOULIAGOS (2010): "Picture This: A Simple Graph That Reveals Much Ado About Research." *Journal of Economic Surveys* **24(1)**: pp. 170–191.

STANLEY, T. D. & H. DOUCOULIAGOS (2011): "Meta-Regression Approximations to Reduce Publication Selection Bias." *Economics Series 2011/4*, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

STANLEY, T. D., H. DOUCOULIAGOS, & S. B. JARRELL (2008): "Meta-regression analysis as the socio-economics of economics research." *The Journal of Socio-Economics* **37(1)**: pp. 276–292.

STANLEY, T. D. & S. B. JARRELL (1989): "Meta-Regression Analysis: A Quantitative Method of Literature Surveys." *Journal of Economic Surveys* **3(2)**: pp. 161–70.

STERNER, T. (1991): "Gasoline demand in the OECD: Choice of model and data set in pooled estimation." *OPEC Review* **91**: pp. 91–101.

STORCHMANN, K. (2005): "Long-Run Gasoline demand for passenger cars: The role of income distribution." *Energy Economics* **27(1)**: pp. 25–58.

SUTTON, A. J., K. R. ABRAMS, D. R. JONES, T. A. SHELDON, & F. SONG (2000): *Methods for Meta-analysis in Medical Research*. Chichester: Wiley.

TISHLER, A. (1983): "The demand for cars and gasoline: A simultaneous approach." *European Economic Review* **20(1-3)**: pp. 271–287.

URI, N. D. & S. A. HASSANEIN (1985): "Testing for stability: Motor gasoline demand and distillate fuel oil demand." *Energy Economics* **7(2)**: pp. 87–92.

WADUD, Z., D. GRAHAM, & R. NOLAND (2009): "A cointegration analysis of gasoline demand in the United States." *Applied Economics* **41(26)**: pp. 3327–3336.

WEST, S. E. & R. C. WILLIAMS III (2007): "Optimal taxation and cross-price effects on labor supply: Estimates of the optimal gas tax." *Journal of Public Economics* **91(3-4)**: pp. 593–617.

Wheaton, W. C. (1982): "The Long-Run Structure of Transportation and Gasoline Demand." *Bell Journal of Economics* **13(2)**: pp. 439–454.

Wirl, F. (1991): "Energy demand and consumer price expectations: An empirical investigation of the consequences from the recent oil price collapse." *Resources and Energy* **13(3)**: pp. 241–262.

Chapter 5

Publication Bias in the Literature on FDI Spillovers

Abstract

In this paper we conduct a large quantitative survey of the literature on horizontal and vertical spillovers from foreign direct investment (FDI). We create a unique database of spillover estimates for each country examined in the literature. Next, we estimate the average effect corrected for publication selection bias (the preferential selection of positive and significant estimates for publication). Our results suggest that an average reported estimate of backward spillovers is statistically significant. Publication selection is evident only among studies published in peer-reviewed journals, and only among the estimates that authors consider most important. Authors with small data sets engage in more publication selection. The intensity of selection in the literature decreases over time, which supports the economics-research-cycle hypothesis.

Keywords:	Foreign direct investment, meta-analysis, productivity spillovers, publication selection bias
JEL Codes:	C83, F23

This paper is a joint work with Tomas Havranek. We are grateful to Joze Damijan, Ziliang L. Deng, Adam Gersl, Galina Hale, Chidambaran Iyer, Molly Lesher, Marcella Nicolini, Pavel Vacek, and Katja Zajc-Kejzar for sending additional data, or explaining the details of their methodology, or both. We thank Beata Javorcik, Tom Stanley, and Pavel Vacek for their helpful comments on an early working-paper version. The paper also benefited from discussions at the Meta-Analysis of Economics Research Colloquium, Conway, 2010; and the Global Development Conference, Bogota, 2011. The paper has been published in the Journal of Development Studies.

5.1 Introduction

Policy makers, especially in transition and developing countries, usually encourage inward FDI in expectation that domestic firms in the same sectors benefit from know-how brought by foreigner investors. Moreover, many of such policy makers believe that firms in supplier sectors benefit from direct knowledge transfers from foreigners, and perhaps also that firms in customer sectors benefit from higher-quality intermediate inputs produced by foreigners. With an allusion to the production chain, the effect of foreign presence on the productivity of domestic competitors is typically labeled *horizontal spillovers*, the effect on domestic suppliers *backward spillovers*, and the effect on domestic customers *forward spillovers*; backward and forward spillovers together are called vertical spillovers. Although not a necessary nor sufficient condition for the provision of government subsidies for FDI, spillover effects are highly policy-relevant. In consequence, the search for spillovers has given rise to a burgeoning stream of empirical literature in development economics, and we investigate 57 such papers in this meta-analysis.

Horizontal spillovers are usually thought to occur through three main channels. The first channel is the competition effect (for example, Aitken & Harrison 1999): the entry of foreign firms increases competition in the domestic market. Increased competition forces domestic firms to use their inputs more efficiently, boosting their productivity. Nonetheless, increased competition also reduces the opportunities of domestic firms to exploit returns to scale, reducing their productivity. The second channel is the demonstration effect (for example, Blomstrom & Kokko 1998): foreign investors bring technology more advanced than that of domestic firms, especially in transition and developing countries. In this way, foreigners "demonstrate" up-to-date technology to domestic firms, which imitate and implement it. The third channel is labor turnover (for example, Görg & Greenaway 2004): foreign firms train local employees, who accumulate know-how and experience with modern technology. Eventually, locals change the employer or start a firm of their own, diffusing knowledge further.

Foreign affiliates will try to prevent the transfer of technology to their competitors; that is, they will try to minimize the positive effects of demonstration and labor turnover. Therefore if the detrimental effects of competition prevail, horizontal spillovers altogether may well become insignificant or even negative. On the other hand, foreigners have incentives to provide assistance to their local suppliers, since they want to ensure a high quality and on-time delivery of inputs.

Indeed, anecdotal evidence indicates that local suppliers may benefit from the interactions with foreign investors, even if investors nor suppliers are particularly knowledge-intensive. In a recent interview conducted by the authors of this paper, the chief executive officer of a Czech printing house describes how the company benefited from the contacts with a Japanese investor: The investor, doing business in electronics, was seeking a local contractor to print millions of instruction manuals for the European market. After the Czech company had won the contract, the representatives of the Japanese investor inspected the company and requested specific improvements in quality management. The representatives had gained experience from their contacts with suppliers in Japan, and they asked no compensation for the advice. The Czech printing house, in turn, applied the improvements also in other areas of production.

Therefore, the recent literature (Javorcik 2004; Blalock & Gertler 2008) emphasizes vertical linkages between foreign investors and domestic firms. The per-job value of spillovers stirred up by linkages can be compared with the amount of government subsidies, as Haskel *et al.* (2007) do; hence for policy recommendations precise estimates of spillovers are required. Since the results of individual studies vary broadly, a quantitative literature survey, meta-analysis, represents a useful method to obtain robust estimates of spillovers (Stanley 2001). If, however, some particular results are more likely to be published (for example, those consistent with the mainstream intuition about spillovers outlined in the paragraphs above), a simple average of the reported results will be a biased estimate of the underlying spillover effect. The importance of publication selection bias in the spillover literature was stressed already by the first meta-analysis on this topic, Görg & Strobl (2001).

Publication bias has been identified in many areas of economics research (Doucouliagos & Stanley 2013). It stems from the preference of authors, editors, or reviewers for some particular results; usually those that are statistically significant or consistent with theory (Stanley 2005). Publication bias can seriously exaggerate the magnitude of the underlying effect, which has been the case, for example, of the negative effects of minimum-wage increases on employment (Doucouliagos & Stanley 2009), the price elasticity of gasoline demand (Havranek *et al.* 2012), or the positive effects of currency unions on trade (Havranek 2010). In a large survey of economics meta-analyses, Doucouliagos & Stanley (2013) find that the magnitude of publication bias decreases with more theory competition in the particular research area. Stanley *et al.* (2008) show that publication selection is a complex phenomenon affected, among other things, by the characteristics of individual researchers. To our knowledge, however, no study has yet systematically examined the micro-level determinants of publication bias.

In contrast to the earlier meta-analyses on FDI spillovers (Görg & Strobl 2001; Meyer & Sinani 2009), we examine backward and forward spillovers in addition to horizontal spillovers. Using a large data set, we employ modern meta-analysis methods developed by Stanley (2005; 2008) to estimate the underlying spillover effects and the magnitude of publication bias. We present individual surveys for each country inspected in the literature and construct a unique cross-country data set of estimated spillovers. Furthermore, we retrieve estimates of publication bias for each study and examine how the intensity of publication selection depends on the characteristics of the authors, such as affiliation, experience, and tenure pressure.

The rest of the paper is structured as follows: Section 5.2 discusses how FDI spillovers are measured, Section 5.3 describes how we extracted information from primary studies, Section 5.4 presents the estimation of publication bias and true spillover effects, Section 5.5 focuses on the determinants of publication bias, and Section 5.6 concludes the paper. The Appendix provides meta-analyses for individual studies and countries.

5.2 Measuring Productivity Spillovers

To estimate the size of productivity spillovers from foreign direct investment, researchers usually examine the relation between foreign presence and the productivity of domestic firms. The variable corresponding to foreign presence is defined depending on the type of spillover under investigation. For horizontal spillovers, foreign presence simply denotes the ratio of foreign activity to the total activity in the domestic firm's own sector. For backward spillovers, foreign presence is defined as the ratio of foreign activity in sectors that buy intermediate goods from the domestic firm. Finally, in the case of forward spillovers foreign presence denotes the ratio of foreign activity in sectors that sell intermediate goods to the domestic firm. Most researchers include all these variables in one regression, together with a number of control variables:

$$\ln \text{Productivity}_{ij} = e_0^h \cdot \text{Horizontal}_j + e_0^b \cdot \text{Backward}_j + e_0^f \cdot \text{Forward}_j + \boldsymbol{\alpha} \cdot \text{Controls}_{ij} + u_{ij}, \tag{5.1}$$

where i denotes domestic firms and j denotes sectors.

Approximately 90% of all studies use firm-level panel data to estimate equation (5.1). Still, a few cross-sectional sector-level studies have been published after year 2000, even though Görg & Strobl (2001) showed that cross-sectional studies systematically overstate horizontal spillovers. Most authors use the share of sector output as a measure of foreign presence, but some use shares of sector employment or equity. The most common control variables include sector competition, demand in downstream sectors, and a measure of absorption capacity (such as the technology gap between domestic and foreign firms or domestic firms' expenditures on research and development).

The majority of authors employ total factor productivity (TFP) as the measure of productivity, while others use output, value added, or labor productivity for the response variable. When computing TFP, most authors take into account the endogeneity of input demand and use the Levinson-Petrin or Olley-Pakes method, but

10% of all estimates are computed using ordinary least squares. Approximately a half of all studies estimate equation (5.1) in differences. A general-method-of-moments estimator is employed by 9% of the studies, and the translog production function instead of the usual Cobb-Douglas function is employed by 8% of them.

Despite these various methods, the results of all studies boil down to the estimates of coefficients e from (5.1), which are directly comparable whenever the log-level specification is used. (The heterogeneity of estimates with respect to different estimation techniques in this literature is explored in detail in a companion article, Havranek & Irsova 2011.) The coefficients represent the semi-elasticity of the productivity of domestic firms with respect to foreign presence:

$$e_0 \approx (\% \text{ change in productivity})/(\text{change in foreign presence}), \quad \text{for. pr.} \in [0, 1]; \tag{5.2}$$

that is, approximately the percentage increase in domestic productivity associated with a one-percentage-point increase in foreign presence. Semi-elasticity has been previously used in meta-analysis, for example, by Rose & Stanley (2005) and Feld & Heckemeyer (2011), and represents the natural choice of summary statistic for the spillover literature.

5.3 Studies on Spillovers from FDI

Because Görg & Strobl (2001) conducted a meta-analysis of horizontal spillover studies published before 2000, in this paper we focus on studies originated after 2000, and especially on the new subset of the literature: vertical spillovers (another reason for focusing on post-2000 studies is that the earlier studies are often not directly comparable). Our intention is to collect all studies estimating backward and forward spillovers; nonetheless, if these studies also estimate horizontal spillovers, we use that information as well. Thus our meta-analysis can be viewed as a complete meta-analysis of vertical spillovers and a partial meta-analysis of horizontal spillovers.

We searched the EconLit, Scopus, Google Scholar, and RePEc databases for prospective studies. Additionally we examined the references of all identified studies published in the last year of our sample, 2010, and also the citations of the most influential article on vertical spillovers, Javorcik (2004). We excluded a few studies that estimated vertical spillovers but that did not define foreign presence as a ratio and hence could not be used to compute semi-elasticity (for example, Kugler 2006; Bitzer *et al.* 2008). Approximately 20% of all studies in our sample could be included only thanks to the cooperation with authors, who often sent us the standard errors of the estimated elasticities. Following the advice of Stanley (2001), "better err on the side of inclusion," we did not exlude any study based on the form, place, or language of publication. The last study was included on 31 March 2010. We do not update studies after that date—for example, if a working paper is included in our data set and gets published after 31 March 2010, we still use the working-paper version.

We collect all estimates reported in the studies. It would be inefficient to discard data; moreover, often it is not clear which estimate the authors prefer. Later in the analysis, as a robustness check, we use averages of all reported coefficients from each study. We gather 57 studies that include 3,626 estimates of elasticity for different types of spillover. Coding such a large number of observations manually is a laborious exercise, and it is difficult to avoid mistakes in the process. To minimize the danger of typos, both of us collected all data independently; it is unlikely that both collectors would make the same mistake. At the end we compared our data sets, reached consensus for every data point that differed, and retrieved the data set.

We provide a summary of all studies in Table 5.6 and Table 5.7 in the Appendix and display representative spillover coefficients for each study. The representative semi-elasticities are estimated as inverse-variance-weighted averages with individual random effects to take into account heterogeneity within studies; the method is called the simple random-effects meta-analysis. Additional details on data properties and collection can be found in the working-paper version of this article, Havranek & Irsova (2010).

Most narrative reviews of empirical literature only consider studies published in high-quality journals. We begin the analysis with a set of such studies to illustrate how the restriction of the sample may, under realistic conditions, lead to biased conclusions concerning the strength of the examined phenomenon. We define high-quality journals for spillover literature as the leading outlets in international economics (Journal of International Economics), international business (Journal of International Business Studies), and development economics (Journal of Development Economics). Naturally, one study published in the American Economic Review is also included in the subset, increasing the number of identified studies to seven. The selected journals have the highest impact factor in the sample, and even if we added the journal with the next highest impact factor (The World Economy), the inference would be similar.

Table 5.1: Qualitative results of studies published in high-quality journals

Study	Journal	Backward	Forward	Horizontal
Javorcik (2004)	American Econ. Rev.	+	?	?
Bwalya (2006)	Journal of Dev. Econ.	+		−
Kugler (2006)	Journal of Dev. Econ.	+[a]	+[a]	?
Blalock & Gertler (2008)	Journal of Int. Econ	+		?
Javorcik & Spatareanu (2008)	Journal of Dev. Econ.	+[b]		−
Liu (2008)	Journal of Dev. Econ.	+[c]	?	+[c]
Blalock & Simon (2009)	Journal of Int. Bus. St.	+		?
Liu *et al.* (2009)	Journal of Int. Bus. St.	+	+	−

Note: +, −, and ? denote the finding of positive, negative, and insignificant spillover effects.

[a] The author does not discriminate between backward and forward spillovers.

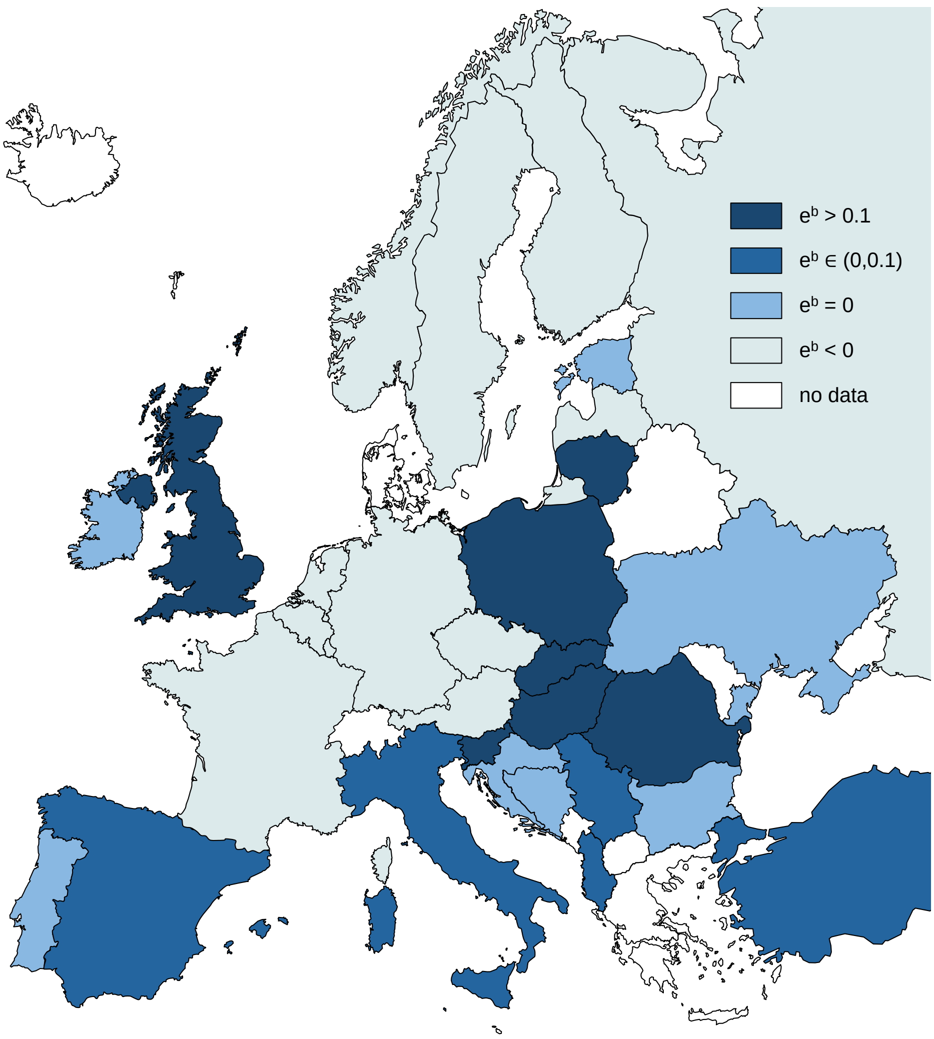
[b] Positive effect reported only for investments with joint foreign and domestic ownership.

[c] Positive long-run effect, negative short-run effect.

Table 5.1 summarizes the qualitative results of studies published in high-quality journals. We add Kugler (2006) to the table since the study is frequently cited in the literature, even if its quantitative results are incomparable with studies in our sample; he does not estimate semi-elasticity, hence we cannot include the study in our quantitative analysis that follows. From Table 5.1 it is apparent that the evidence for positive and significant backward spillovers is unequivocal, but no such consensus emerges for forward and horizontal spillovers: some researchers report positive effects of forward linkages and negative effects of horizontal linkages; others find insignificant effects.

Taking a simple average of all estimates reported in high-quality journals confirms this qualitative observation. The average semi-elasticity reaches 1.14 for backward spillovers, 0.54 for forward spillovers, and −0.13 for horizontal spillovers, all significant at the 5% level. Most of the studies concentrate on backward spillovers and provide estimates of forward and horizontal spillovers only as a bonus. The practice reflects the recent view that domestic firms supplying foreign affiliates are the most likely beneficiaries of technology transfer and that the effect on competitors and customers is less important.

Figure 5.1: Cross-country evidence on backward spillovers



It is remarkable that the average estimates of spillovers do not change significantly if we broaden the sample to all published studies, but the averages shrink a lot if unpublished studies are considered as well. If unpublished studies are included, the average backward spillover reaches only 0.27, the average forward spillover is negative (-0.09) and insignificant, and the average horizontal spillover is negligible (0.01). All of this points to selection bias in published studies, which will be the topic of Section 5.4.

Because the studies in our sample estimate spillover coefficients for many different countries, they allow us to construct a unique cross-country database of spillovers from foreign direct investment. The database is summarized in the Appendix in Table 5.8; we use the simple random-effects meta-analysis to estimate the average backward, forward, and horizontal spillover based on the entire available literature for each country. The results for European countries are depicted in Figure 5.1 (the graphical presentation of our results is not convenient for other continents as some large countries have not yet been studied in the spillover literature). It is apparent that the effects of backward linkages are relatively heterogeneous. While it is difficult to draw general conclusions, the figure, in line with Bitzer *et al.* (2008), suggests that Central-Eastern European countries may benefit relatively more from foreign investment than their advanced western-European counterparts. The average backward spillovers, estimated by the simple random-effects meta-analysis, also significantly differ between the income categories of countries in our sample: spillovers are larger for developing economies (0.22) than for developed economies (0.09). The cross-country heterogeneity in the sample is examined in detail in a companion article (Havranek & Irsova 2011); in this paper we focus on publication bias.

5.4 Quantifying Publication Bias

Publication selection bias in the FDI spillover literature was first examined by the well-known meta-analysis of Görg & Strobl (2001). Following Card & Krueger (1995), they used two distinct tests of publication bias, and both tests indicated the presence of publication bias in the literature. Nevertheless, the number of observations available to Görg & Strobl (2001) for the tests was only 23 and 16, respectively. We revisit the issue of publication bias in this broad literature taking the advantage of three times more new primary studies published after year 2000 and more than 1,000 estimates for each type of spillover.

The first test of publication bias draws on Begg & Berlin (1988). It examines the relation between the reported t-statistic of the spillover coefficient and the number of degrees of freedom available for the estimation. Görg & Strobl (2001) argue that, in the absence of publication bias, the absolute value of t-statistic should increase with more degrees of freedom (roughly speaking, more observations make the estimation more precise and increase statistical significance). Specifically, Card & Krueger (1995) explains that the logarithm of the absolute value of the t-statistic should be directly proportional to the logarithm of the square root of the number of degrees of freedom. As Stanley (2005) notes, it makes no practical difference whether the number of observations or degrees of freedom is used for this test, and because spillover studies directly report the number of observations, we employ the following specification:

$$\log|t_i| = \alpha_1 + \alpha_0 \log(\sqrt{\text{Number of observations}})_i + \epsilon_i, \tag{5.3}$$

where $\alpha_0 = 1$ if no publication bias is present. The specification is estimated separately for each type of spillover, and the results are summarized in Table 5.2. Similarly to Görg & Strobl (2001), we reject the hypothesis $\alpha_0 = 1$; the result of the test is the same for horizontal, backward, and forward spillovers.

Table 5.2: Test of publication bias following Görg & Strobl (2001)

	Backward	Forward	Horizontal
$\log\sqrt{\text{number of observations}}$	0.112^{***}	0.512^{***}	0.206^{***}
	(0.0355)	(0.0446)	(0.0400)
Constant	-0.253	-2.187^{***}	-0.832^{***}
	(0.168)	(0.215)	(0.197)
Observations	1401	1067	1204
Studies	56	45	52
t-stat (H_0: $\alpha_0 = 1$)	25.0^{***}	11.0^{***}	19.9^{***}

Note: Estimated by OLS; heteroscedasticity-robust standard errors in parentheses.
Response variable: logarithm of the absolute value of t-statistic of the estimate of semi-elasticity.
*** denotes significance at the 1% level.

But does it really have to mean that we have found evidence for publication selection bias? Stanley (2005) and Doucouliagos & Stanley (2009) show that specification (5.3), emphasized by both Card & Krueger (1995) and Görg & Strobl (2001), is not a proper one to test for publication bias. For example, consider the case when there is no genuine empirical effect. Since here we have $t = e/SE$ and $e = 0$, the absolute value of t-statistic does not increase with more observations, even though more observations reduce standard errors. Hence, if a weak relation between the t-statistic and the number of observation is found, it may suggest either the presence of publication bias or the absence of the underlying empirical effect.

Stanley (2005) argues that specification (5.3) should be interpreted as a test for genuine empirical effect. Note that the relationship between the absolute value of t-statistic and the number of observations is always positive and significant in Table 5.2, which may indicate that the underlying spillover effects are different from zero. Stanley (2005), however, warns that this test has large type I errors (false rejection of no relationship), especially if the literature suffers from misspecification biases. Since the meta-analysis of Görg & Strobl (2001) suggests that misspecifications indeed drive some results in this literature (specifically, they find that studies using cross-sectional data overstate spillover effects), we take the evidence from Table 5.2 with a grain of salt.

Instead, the second test presented by Card & Krueger (1995) and Görg & Strobl (2001) has become the cornerstone of modern meta-analysis, and can be used to detect the significance and magnitude of both publication bias and genuine underlying effect (Egger *et al.* 1997; Stanley 2005; 2008). Before turning to formal regression analysis, it is beneficial to introduce the graphical version of this test (Stanley & Doucouliagos 2010). The so-called funnel plot is the most common method of detecting publication bias in medical meta-analyses (Sutton *et al.* 2000). It is a scatter plot of the size of the estimates on the horizontal axis and their precision, usually the inverse of standard errors, on the vertical axis. The most precise estimates will be close to the genuine underlying effect (that is, the top portion of the scatter should be narrow), while imprecise estimates will be more dispersed (that is, the bottom portion should be wide). In sum, the cloud of the estimates should resemble an inverted funnel. Publication bias is then indicated by the asymmetry of the funnel plot: in the absence of publication bias, all imprecise estimates have the same chance of being reported, and the funnel is symmetric.

The funnel plots constructed using estimates from both published and unpublished studies for the three types of spillover are depicted in the top panels of Figure 5.2. One thing that strikes an economics meta-analyst is the general symmetry of all three funnels: a relatively rare sight considering that the meta-meta-analysis of Doucouliagos & Stanley (2013) finds "substantial" or "serious" publication bias for two thirds of all areas in empirical economics. Close inspection of the funnels, however, reveal slight asymmetries. The right-hand part of the funnel for backward spillovers is a little heavier, suggesting possible selection in favor of positive estimates of backward spillovers. On the other hand, for horizontal spillovers the funnel is slightly skewed to the left, which indicates possible preferential selection of negative estimates. We can identify no sign of bias for forward spillovers, although the funnel is worse-shaped compared with the other two: perhaps there is greater heterogeneity among forward spillovers, which causes some precise estimates to lie far away from the mean value.

Despite its name, publication selection in economics usually does not restrict to published studies. Rational authors are likely to polish even early drafts of their papers if they expect that some particular results are more likely to be accepted for publication (or they use a particular direction of estimates as a specification check and discard estimates with an unintuitive sign). For this reason, most meta-analysts pool published and unpublished papers together when testing for selection bias. If the pattern of publication bias was stronger for published than for unpublished studies, we would have a reason to believe that, aside from self-censorship, there is an additional selection pressure from journal editors and reviewers.

The bottom panels of Figure 5.2 show funnel plots for the three types of spillover when only published studies are considered. It is apparent at first sight that the slight asymmetries identified earlier in the funnels for all estimates now become much more

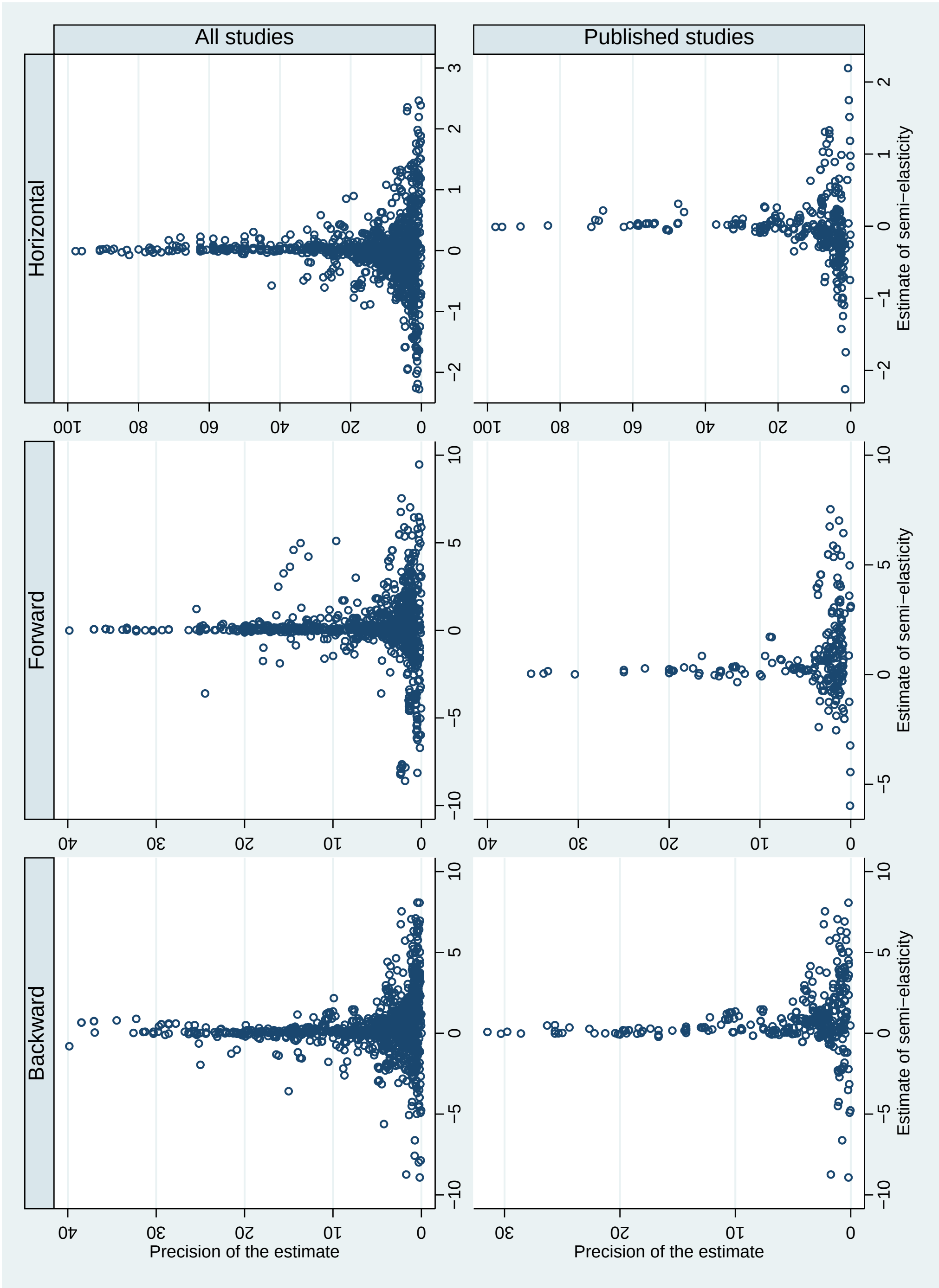


Figure 5.2: Funnel plots show publication bias in journal articles

pronounced. Clear biases emerge, upward for backward spillovers and downward for horizontal spillovers. Such a different direction of publication selection of related coefficients taken from the same literature is surprising.

When the first influential article on backward spillovers was published (Javorcik 2004), the article in the American Economic Review stressed the contrast between its findings of large positive backward spillovers and negligible horizontal spillovers. Since the consensus at that time was that the evidence for horizontal spillovers was mixed at best (Görg & Greenaway 2004), Javorcik (2004) argued that researchers "have been looking for spillovers in the wrong place." This appealing argument has been repeated many times in the burgeoning literature that has followed Javorcik (2004) and that has estimated backward spillovers for different countries. Now, if authors used this result as a specification check, upward selection bias for backward spillovers and downward bias for horizontal spillovers would follow. This is precisely what the funnel plots suggest.

The interpretation of funnel plots, however, is subjective; asymmetry of the funnel is difficult to detect precisely by mere visual inspection. Thus a formal version of the test for asymmetry is necessary. It follows from rotating the axes of the funnel plot, so that the effect size is now on the vertical axis, and from inverting the values of precision on the new horizontal axis. The inverted scatter plot can be interpreted as a regression relationship (Card & Krueger 1995; Görg & Strobl 2001):

$$e_i = e_0 + \beta_0 \cdot Se(e_i) + u_i, \tag{5.4}$$

where β_0 measures the asymmetry of the funnel plot and the strength of publication bias. In the absence of publication bias we would not expect the reported coefficients to be correlated with their standard errors, and the estimates would be randomly distributed around the true effect e_0. Because equation (5.4) is heteroscedastic by definition, and the measure of heteroscedasticity is Se (the explanatory variable is a sample estimate of the standard deviation of the response variable; the heteroscedasticity is also apparent from the funnel plots), the specification is usually estimated by weighted least squares (Stanley 2005):

$$e_i/Se(e_i) = t_i = e_0 \cdot 1/Se(e_i) + \beta_0 + \xi_i. \tag{5.5}$$

Though simple, the specification is remarkably efficient for the estimation of both the magnitude of publication bias (β_0) and the true effect (e_0), as Monte Carlo simulations show (Stanley 2008). Because we have a large number of observations, and results within studies are likely to be correlated, we include study dummies in the estimation to purge away individual study effects.[1] We use heteroscedasticity-robust standard errors clustered at the study level. As a robustness check, we also use only one estimate per each study; the representative estimate for each study is estimated by the simple random-effects meta-analysis. To account for possible outliers, we additionally run the regression with one estimate per study using a robust MM-estimator introduced by Verardi & Croux (2009), who show that this modern method outperforms all other commonly used robust estimators. The results for all three types of spillover are summarized in Table 5.3.

When both published and unpublished studies are considered together, Table 5.3 shows hardly any evidence of bias for any type of spillover. Only the coefficient for forward spillovers is significant at the 10% level, but the magnitude is small (Doucouliagos & Stanley 2013, would label it as "little to modest") and the significance disappears in the robustness check. Thus formal meta-regression methods do not corroborate the slight asymmetries identified in the top three funnel plots of Figure 5.2. If, on the other hand, only published studies are considered, the asymmetry of the funnel plot for backward spillovers is corroborated strongly: on average, published studies significantly overstate the magnitude of backward spillovers. The meta-regression does not support the asymmetry of the funnel plot for horizontal spillovers, however.

[1] Some confusion exists in the meta-analysis literature concerning the term "fixed-effects estimation." Sometimes a simple OLS estimation is labeled "fixed effects" to emphasize that no random-effect component is present; the traditional fixed-effects estimation, on the other hand, is rarely applied in meta-analysis since studies reporting only one estimate are dropped in the procedure (see, for instance, Nelson & Kennedy 2009). In our case, though, the large data set of spillover estimates allows us to use the traditional fixed-effects estimation, which is less restrictive than random effects.

Table 5.3: Test of publication bias and true effect following Stanley (2005)

	Fixed effects		Robust
Backward Spillovers	All	Published	All
Constant (publication bias)	-0.102	1.760***	1.509
	(0.315)	(0.0426)	(1.038)
$1/Se$ (true effect beyond bias)	0.124***	0.0934***	0.0371*
	(0.0383)	(0.00519)	(0.0188)
Observations	1402	401	56
Studies	56	27	56
	Fixed effects		Robust
Forward Spillovers	All	Published	All
Constant (publication bias)	0.940*	-0.160	-0.287
	(0.524)	(0.758)	(0.710)
$1/Se$ (true effect beyond bias)	-0.0388	0.342**	0.0294***
	(0.0567)	(0.123)	(0.00960)
Observations	1066	262	45
Studies	45	20	45
	Fixed effects		Robust
Horizontal Spillovers	All	Published	All
Constant (publication bias)	-0.312	0.0750	0.800
	(0.261)	(0.0446)	(0.784)
$1/Se$ (true effect beyond bias)	0.0214	0.00649***	0.00624
	(0.0147)	(0.00230)	(0.00739)
Observations	1205	352	52
Studies	52	27	52

Note: Heteroscedasticity-robust standard errors in parentheses. Response variable: t-statistic of the estimate of semi-elasticity.
Fixed effects = Study fixed effects are included. Robust = the simple random-effects meta-analysis is run for each study separately; then, using a robust MM-estimator following Verardi & Croux (2009), the meta-regression is run on the results. All = estimates from all studies. Published = only estimates from studies published in refereed journals.
***, **, and * denote significance at the 1%, 5%, and 10% levels.

In sum, we found formal evidence of publication selection only for backward spillovers, and only for published studies. No selection was found for the estimates of forward and horizontal spillovers and for unpublished estimates of any type of spillover. In the spillover literature, horizontal and forward spillovers are generally viewed as less important than backward spillovers: for forward spillovers it is relatively difficult to find interpretation, and horizontal spillovers are often expected to be insignificant due to the competition effect (Görg & Greenaway 2004). In the post-2000 literature, which constitutes our sample, the estimates of backward spillovers make the story of the paper; hence, they are more likely to be polished.

When estimating the genuine spillover effects after correction for within-study correlation and publication bias, it is meaningful to restrict our interpretation to the specifications that include all studies (if only published studies were considered, at least 70% of all observations would be discarded; we have also found that published studies exhibit more bias). Both estimates of the true effect for backward spillovers are positive and statistically significant at least at the 10% level. We prefer the fixed-effects model, where information from all 1,402 observations is used: in this case, the estimate of backward spillover is significant at the 1% level and reaches 0.12. In other words, a 10-percentage-point increase in foreign presence is on average associated with an incerase in the productivity of local firms in supplier sectors by 1.2%, an economically significant effect.

The results for forward spillovers are mixed: the fixed-effects estimator reports an insignificant coefficient, while the robust estimator reports a coefficient that is highly significant (though small in magnitude). In either case, forward spillovers do not seem to be economically important. Finally, for horizontal spillovers both estimators show insignificant results.

As Stanley & Doucouliagos (2007) note, the estimates of e_0 from regression (5.5) may be biased downwards, and they recommend estimating the Heckman meta-regression, which assumes a quadratic relation between estimated effects and their standard errors in equation (5.4). But because in our case no publication selection is present in the sample of all studies, the coefficient estimated using the Heckman regression is very close to that estimated by (5.5), and so we do not report it.

5.5 Determinants of Publication Selection

As funnel plots depicted in Figure 5.2 show, the pattern of publication bias can vary significantly among subsets of studies and among related topics within a single empirical literature. In this section we aim to explore the sources of this heterogeneity and shed some light on the determinants of publication selection. We concentrate on backward spillovers, because we have already found evidence of a large difference in

Table 5.4: Summary statistics of explanatory variables

Variable	Mean	Std. dev.	Min	Max
Study was published in a peer-reviewed journal	0.482	0.504	0	1
Journal or series impact factor	0.194	0.431	0	2.70
Number of citations of the most cited co-author (logarithm)	2.83	2.34	0	6.66
Number of observations used by the study (logarithm)	9.40	2.12	3.76	13.7
A co-author is native to the country under examination	0.723	0.444	0	1
A co-author is affiliated with a US-based institution	0.250	0.437	0	1
A co-author is affiliated with an academic institution	0.643	0.483	0	1
Year and month when the study was published (base: 2000)	7.77	1.73	2.79	10.5
Focus on the interpretation of spillover significance	0.536	0.503	0	1
The corresponding author has not completed PhD	0.268	0.447	0	1
The corresponding author has completed PhD 1–5 years ago	0.304	0.464	0	1
The corresponding author has completed PhD 6–10 years ago	0.250	0.437	0	1

Note: Data on the number of citations and the impact factor (recursive) are taken from RePEc.

publication bias between published and unpublished studies in this area. The topic of this section is related to the meta-meta-analysis by Doucouliagos & Stanley (2013), who study the determinants of publication selection *across economics literatures*—that is, on the macro level. We, instead, focus on of the individual studies' and authors' characteristics that may influence the strength of publication bias within one literature.

As the response variable, we employ study-level estimates of the extent of publication bias taken from regression (5.5). Because the intercept in (5.5) is a measure of publication bias, and the specification was estimated with fixed effects for individual studies, we use the estimated fixed effects to compute individual intercepts. These intercepts, in turn, measure the extent of publication bias for each study. A large data set is needed for such an exercise. Fortunately, we have 56 studies that provide more than one estimate of backward spillovers, and the vast majority of them provide more than 10 estimates. In any case, however, some of the individual intercepts will not be estimated precisely, and our response variable will therefore have a relatively large random sampling error. The estimation is likely to produce several outlying values of publication bias, especially for studies that provide only a few estimates of backward spillovers. For this reason, we employ a robust MM-estimator introduced by Verardi & Croux (2009) in all following regressions.

We select a dozen study-level variables that may potentially influence the intensity of publication selection in favor of significant positive estimates of backward spillovers; the summary statistics of these variables are reported in Table 5.4. First, we include a dummy variable that equals one if the study was published in a peer-reviewed journal: as shown in Section 5.4, we expect that published studies exhibit more publication selection. Approximately half of the studies were published in journals, the rest of our observations were obtained from working papers and dissertations. Another explanatory variable is the impact factor of the journal or working-paper series: perhaps the pattern of publication bias differs between lower- and higher-quality outlets.

If the selection of some particular results increases the probability of publication, successful authors will have more experience with polishing (either intentional or unintentional), and we expect greater publication bias in studies they co-author. A fine measure of authors' success is the number of citations, and we include it as an additional explanatory variable (for each author we collect all RePEc citations received before the date of publication of the particular study). Next, because publication bias should be inversely related to sample size (Stanley 2005), we expect that studies with larger data sets will be engaged in less publication selection. As shown in Section 5.4, the underlying average backward spillover is positive; therefore authors with many observations will not have to search long for intuitive (that is, positive and significant) results.

We include a dummy which equals one if at least one co-author of the study is native to the country under investigation, which holds for 72% of studies in our sample. We consider an author native if he either was born in the country or obtained an academic degree there. We expect that such authors may have vested interest in the results and may be involved more in publication selection. Another included dummy equals one if a co-author is affiliated with a US-based institution (25% of all studies; we collect affiliations stated by the authors in the pdf versions of their studies). Because of the characteristics of the tenure system in the USA, such authors are likely to be under fierce pressure to publish. When this is the case, they may also be more tempted to polish their studies in order to increase the probability of publication. Testing a weaker hypothesis, Stanley (2005) shows that studies using US data exhibit more publication bias. Of course, in general any affiliation with an academic institution may stimulate publication selection because of the requirements for tenure, and we include a corresponding dummy variable.

Stanley *et al.* (2008), building on the work of Goldfarb (1995), discuss the issue of fashion and novelty in economics research: the so-called economics-research-cycle hypothesis. According to the hypothesis, studies in a newly emerged empirical area produce large and significant estimates of the underlying effect at first, but as the time passes, skeptical results become more interesting and soon begin to dominate the literature. Thus we should observe a downward trend in the reported t-statistics. If the underlying effect does not change in time, the downward trend in the reported results is entirely due to publication selection, and therefore implies a similar trend in the extent of publication bias. To test the hypothesis, we add a variable indicating the year and month of study publication.

We expect studies focusing specifically on the interpretation of the size and statistical significance of backward spillovers to exhibit more publication bias: because the most important results make the story of the paper, they are most likely to be polished. On the other hand, some studies estimate FDI spillover regressions, but concentrate on the heterogeneity of spillover effects and not on the magnitude of spillover effects per se (for example, they add interactions of foreign presence with the absorption capacity of domestic firms); we do not expect these studies to involve in much selection of spillover coefficients based on polarity or significance.

Finally, we include dummy variables reflecting the PhD vintage of the corresponding authors of the studies in our sample. Authors are divided into four groups: those who have not completed PhD at the time when their study was published (27% of all studies), those who have completed PhD between 1 and 5 years ago (30%), those who have completed PhD between 6 and 10 years ago (25%), and those who have completed PhD more than 10 years ago (18%). Because the tenure pressure likely magnifies the selection bias, we expect stronger selection among studies published by researchers who have completed PhD less than 6 years ago.

Table 5.5: Determinants of publication selection toward large positive estimates

Explanatory variable	(1)	(2)	(3)
Study was published in a peer-reviewed journal	0.978***	1.028***	0.904**
	(0.284)	(0.304)	(0.337)
Journal or series impact factor	-1.578***	-1.564***	-1.631***
	(0.224)	(0.265)	(0.303)
Number of citations of the most cited co-author	0.192***	0.224***	0.228***
	(0.0612)	(0.0805)	(0.0605)
Number of observations used by the study	-0.362***	-0.354***	-0.362***
	(0.0999)	(0.103)	(0.129)
A co-author is native to the country under examination	0.101	0.0879	
	(0.430)	(0.457)	
A co-author is affiliated with a US-based institution	1.041***	1.031***	0.959**
	(0.365)	(0.371)	(0.401)
A co-author is affiliated with an academic institution	-0.693**	-0.708	
	(0.286)	(0.446)	
Year and month when the study was published	-0.350***	-0.330**	-0.294*
	(0.113)	(0.124)	(0.159)
Focus on the interpretation of spillover significance	0.719**	0.729**	0.710**
	(0.298)	(0.305)	(0.289)
The corresponding author has not completed PhD		0.356	
		(0.528)	
The corresponding author has completed PhD 1–5 years ago		0.211	
		(0.484)	
The corresponding author has completed PhD 6–10 years ago		0.0522	
		(0.675)	
Constant	4.986***	4.471***	4.159**
	(1.335)	(1.656)	(1.555)
Observations	56	56	56
Pseudo R-squared	0.321	0.330	0.306
F-stat (H_0: all coefficients for PhD vintage dummies are zero)		0.170	

Note: Standard errors in parentheses. Response variable: the magnitude of publication bias in the study. Estimated using a robust MM-estimator following Verardi & Croux (2009).
***, **, and * denote significance at the 1%, 5%, and 10% levels.

Regression results are reported in Table 5.5. In the first column we include all explanatory variables except for PhD vintage dummies, which we add in the second column. The third column reports the results of the general-to-specific modeling approach: all insignificant variables are gradually excluded from the model. The R-squared of the regressions varies from 0.31 to 0.33. Because the data we use are in essence micro-level, such relatively low values of R-squared are not surprising; they are common in meta-analysis (see, for example, Disdier & Head 2008). Moreover, as noted earlier, the response variable has a relatively large random sampling error, which means that a part of variation in the response variable cannot be explained by any regression model.

As expected, we find that publication in a peer-reviewed journal is associated with more selection bias. The results are in line with the intuition and evidence presented in Section 5.4, and imply that, other things equal, selection bias in published studies is larger by 1 compared to unpublished studies. The difference is enough to move the incidence of publication selection from one category listed by Doucouliagos & Stanley (2013) to another: for example, from "lesser to moderate" to "substantial." To be more specific, since the average standard error among the estimates of backward spillovers reaches 1.4, from equation (5.4) it follows that published studies exaggerate backward spillovers by the same amount (1×1.4). Because the true underlying spillover is only 0.12, the exaggeration is more than tenfold (recall the average 1.14 from published studies reported in Section 5.3).

Our results indicate that studies published in better outlets (with a higher impact factor) are associated with less publication bias. This could be the case if especially lower-ranked journals used the finding of positive spillovers as an intuitive check of correct specification; on the contrary, high-quality journals are more likely to select studies according to their methodology and the rigor of analysis, giving less weight to intuition. Next, we find that frequently cited authors produce studies with greater publication bias. The evidence is in accordance with the intuition given above: successful authors have more experience with polishing their results. Likewise, it is not surprising that studies with larger data sets show less publication bias.

We find no evidence that authors native to the country examined in their studies would be involved in more selection compared with other researchers. On the other hand, we find substantially higher publication bias for US-affiliated authors; this result corroborates the findings of Stanley (2005). The coefficient estimated for the dummy variable for affiliation with an academic institution is negative, which is not consistent with our expectations. The coefficient, however, is statistically significant only in one out of three specifications.

Our results suggest a clear downward trend in publication bias, which is consistent with the economics-research-cycle hypothesis: studies published in early 2000s may have overstated the average backward spillover; in recent years, however, much lower estimates have been reported. Studies focusing on the interpretation of the size and significance of the spillover coefficient exhibit more publication bias, which corresponds with our expectations. The coefficient reaches approximately 0.72, which indicates that such studies on average exaggerate the estimate of backward spillovers by 1 (0.72×1.4). Finally, the dummies for PhD vintage of the corresponding authors are jointly insignificant, which means that, aside from affiliation with US-based institutions, our data show no further pattern of publication selection connected to tenure pressure. We also tried to include interactions of the dummy for US affiliation and dummies for PhD vintage; nevertheless, the interactions were not significant as well (this additional specification is available on request).

5.6 Conclusion

The principal economic argument for the provision of subsidies for FDI is the assumed knowledge spillovers flowing from foreign affiliates to domestic firms. Because of the importance of FDI spillovers for economic policy, a vast body of empirical literature has attempted to quantify these effects. Nevertheless, the results of individual studies vary significantly (for example, depending on the exact method chosen to estimate the spillover effect), making it difficult for policy makers to draw conclusions from the literature. One possibility is to focus solely on studies published in the most respected journals and discard all other outcomes. Studies published in respected journals may be expected to employ better methods than other studies, but altogether they only provide evidence for a few countries. Moreover, the selectivity of top journals may cause that strong (or, in other words, statistically significant) results have a higher probability of publication, which would distort inference.

The distortion of reported results due to publication pressures is called publication selection bias, and it has been found strong in many areas of economics research (Doucouliagos & Stanley 2013). Goldfarb (1995) formulates the economics-research-cycle hypothesis, which has been corroborated empirically for some fields of applied economics by Stanley *et al.* (2008) and Havranek (2010). According to this hypothesis, seminal contributions in applied economics (that is, papers which are the first to estimate a particular effect) tend to report large and significant estimates. Only strong results convince the editors, overcome the barriers to entry, and a new empirical field is born. The large estimates are often corroborated by subsequent research—but, as the time passes, skeptical results become preferred, since they are considered more interesting by the editors, reviewers, and the readership in general. Because of the possibility of publication bias and the research cycle in the spillover literature, we prefer to evaluate a broad sample of empirical studies, making use of the work of dozens development researchers. Moreover, in contrast to cherry picking, inference derived from the entire literature does not depend on any particular methodology employed by the primary study to estimate spillovers.

We gather 3,626 estimates from 57 studies that focus on vertical spillovers—that is, the effect of foreign presence on domestic firms in supplier or customer sectors. Apart from a complete survey of vertical spillovers, we also conduct a partial survey of horizontal spillovers (the effect of foreign investors on domestic firms in the same sectors) by including only those coefficients that researchers estimate in the same regression with vertical spillovers. We employ modern meta-analysis methods to uncover the underlying economic effect of FDI on the productivity of domestic firms.

Our results suggest that the average effect of foreign affiliates on the productivity of their local competitors (horizontal spillover) is economically insignificant. The effect of foreign affiliates on their local customers (forward spillover) is likewise negligible. On the other hand, we detect a statistically significant and economically meaningful effect of foreign affiliates on their local suppliers (backward spillover). Specifically, a 10-percentage-point increase in foreign presence is associated with a 1.2% boost to the productivity of domestic firms in supplier sectors. Such a spillover effect is consistent with subsidies for FDI. Nevertheless, policy makers should exercise caution because the estimates capture more than externalities: studies on FDI spillovers do not account for possible compensations for the transfer of technology (Keller 2009). An exception is Blalock & Gertler (2008), who examine the influence of foreign presence on the profits of Indonesian firms and confirm the positive externality.

While the average backward spillover is robustly positive, it differs significantly across countries. For example, the effect for all developing countries examined by the studies in our sample is twice as large as the average spillover reported for developed countries. The degree of economic development plays an important role in explaining the difference, but it is not the only one. In a companion paper (Havranek & Irsova 2011) we examine in detail what causes the differences in the reported FDI spillovers. We find that both the characteristics of the host country and the characteristics of FDI matter. For example, a larger technology gap of domestic firms with respect to foreign investors is associated with less spillovers. On the other hand, a higher degree of trade openness is associated with more spillovers from inward FDI. The mode of entry of FDI is also important: fully foreign-owned investments generate less positive spillovers than joint projects of foreign and domestic firms. In the present paper we take stock of the empirical research on FDI spillovers and provide a unique database of average estimates for each country examined in the literature.

Remarkably, we find no evidence of publication selection bias for any type of spillover when both published and unpublished studies are considered together. When only published studies are included, we detect substantial upward bias for backward spillovers, but no bias for horizontal and forward spillovers. Because the recent literature considers backward spillovers the most important spillover type, the results concerning backward spillovers are more likely to be polished. Moreover, theory diversity is lower for backward spillovers than for horizontal spillovers. If the competition effect of increased foreign presence gets negative and outweighs the effects of demonstration and labor turnover, horizontal spillovers altogether may well turn negative; on the contrary, no generally accepted theory exists that would allow for negative backward spillovers. Hence, in line with Doucouliagos & Stanley (2013), we find more publication bias in areas with less theory diversity.

Using the estimated meta-regression we retrieve the magnitude of publication bias for each study. We find that publication selection is stronger for studies that are co-authored by successful authors, where success is measured by the number of citations. Furthermore, authors affiliated with US-based institutions appear to engage in more publication selection, possibly because of a higher pressure to publish. On the other hand, studies with larger data sets exhibit less publication bias, because they are likely to detect statistically significant results without much specification search. The extent of publication bias gradually decreases over time, which is consistent with the economics-research-cycle hypothesis. Finally, we find greater publication selection bias for studies that concentrate on the significance and magnitude of spillover coefficients.

In recent years, the methodology employed by studies estimating FDI spillovers has converged to what may be called "best practice;" at least given the existing data. The standard is to employ firm-level data (in contrast to data aggregated at the industry level, which were often employed in the past), use total factor productivity as the response variable and compute it by a method that takes into account the endogeneity of input demand (in contrast to ordinary least squares), estimate the resulting regression in differences, and control for the characteristics of firms (for example, the absorptive capacity) and industries (for example, the degree of competition). Larger data sets, available in recent years especially for emerging economies such as China, has allowed for more precise estimation and, according to our results, also helped reduce publication bias. In our view, the most important avenue for future research in this field is the examination of spillover determinants, both at the firm and country level. Our meta-analysis suggests that backward spillovers are positive and relatively large on average, but also that they differ significantly across countries. Once the sources of these differences are robustly identified, the literature may provide policy makers valuable guidance on which investors are the most beneficial to attract.

References

Aitken, B. J. & A. E. Harrison (1999): "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela." *American Economic Review*

89(3): pp. 605–618.

Atallah Murra, S. (2006): "Revaluando la transmision de spillovers de la IED: Un estudio de productividad para Colombia." *Revista Desarrollo y Sociedad* **57(1)**: pp. 163–213.

Barrios, S., H. Görg, & E. Strobl (2009): "Spillovers Through Backward Linkages from Multinationals: Measurement Matters!" *IZA Discussion Papers 4477*, Institute for the Study of Labor.

Begg, C. B. & J. A. Berlin (1988): "Publication bias: a problem in interpreting medical data." *Journal of Royal Statistical Society A* **151**: pp. 419–463.

Bitzer, J., I. Geishecker, & H. Görg (2008): "Productivity spillovers through vertical linkages: Evidence from 17 OECD countries." *Economics Letters* **99(2)**: pp. 328–331.

Békés, G., J. Kleinert, & F. Toubal (2009): "Spillovers from Multinationals to Heterogeneous Domestic Firms: Evidence from Hungary." *The World Economy* **32(10)**: pp. 1408–1433.

Blake, A., Z. Deng, & R. Falvey (2009): "How does the productivity of foreign direct investment spill over to local firms in Chinese manufacturing?" *Journal of Chinese Economic and Business Studies* **7(2)**: pp. 183–197.

Blalock, G. & P. J. Gertler (2008): "Welfare gains from Foreign Direct Investment through technology transfer to local suppliers." *Journal of International Economics* **74(2)**: pp. 402–421.

Blalock, G. & D. H. Simon (2009): "Do all firms benefit equally from downstream FDI? The moderating effect of local suppliers' capabilities on productivity gains." *Journal of International Business Studies* **40(7)**: pp. 1095–1112.

Blomstrom, M. & A. Kokko (1998): "Multinational Corporations and Spillovers." *Journal of Economic Surveys* **12(3)**: pp. 247–77.

Blyde, J., M. Kugler, & E. Stein (2004): "Exporting vs. Outsourcing by MNC Subsidiaries: Which Determines FDI Spillovers?" *Discussion Paper Series In*

Economics And Econometrics 0411, Economics Division, School of Social Sciences, University of Southampton.

BWALYA, S. M. (2006): "Foreign direct investment and technology spillovers: Evidence from panel data analysis of manufacturing firms in Zambia." *Journal of Development Economics* **81(2)**: pp. 514–526.

CARD, D. & A. B. KRUEGER (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review* **85(2)**: pp. 238–43.

CHANG, S. J., J. CHUNG, & D. XU (2007): "FDI and Technology Spillovers in China." *CEI Working Paper Series 2007-7*, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.

CRESPO, N., M. P. FONTOURA, & I. PROENÇA (2009): "FDI spillovers at regional level: Evidence from Portugal." *Papers in Regional Science* **88(3)**: pp. 591–607.

DAMIJAN, J. P., M. KNELL, B. MAJCEN, & M. ROJEC (2003): "Technology Transfer through FDI in Top-10 Transition Countries: How Important are Direct Effects, Horizontal and Vertical Spillovers?" *William Davidson Institute Working Papers Series 549*, William Davidson Institute at the University of Michigan, Stephen M. Ross Business School.

DAMIJAN, J. P., M. ROJEC, B. MAJCEN, & M. KNELL (2008): "Impact of Firm Heterogeneity on Direct and Spillover Effects of FDI: Micro Evidence from Ten Transition Countries." *LICOS Discussion Papers 21808*, LICOS—Centre for Institutions and Economic Performance, K. U. Leuven.

DISDIER, A.-C. & K. HEAD (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics* **90(1)**: pp. 37–48.

DOUCOULIAGOS, H. & T. D. STANLEY (2009): "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* **47(2)**: pp. 406–428.

DOUCOULIAGOS, H. & T. D. STANLEY (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys*

27(2): pp. 316–339.

Egger, M., G. D. Smith, M. Scheider, & C. Minder (1997): "Bias in meta-analysis detected by a simple, graphical test." *British Medical Journal* **316**: pp. 320–246.

Feld, L. P. & J. H. Heckemeyer (2011): "FDI and Taxation: A Meta-Study." *Journal of Economic Surveys* **25(2)**: p. 233–272.

Fernandes, A. M. & C. Paunov (2008): "Foreign direct investment in services and manufacturing productivity growth: evidence for Chile." *Policy Research Working Paper Series 4730*, The World Bank.

Gersl, A. (2008): "Productivity, Export Performance, and Financing of the Czech Corporate Sector: The Effects of Foreign Direct Investment." *Czech Journal of Economics and Finance* **58(05-06)**: pp. 232–247.

Gersl, A., I. Rubene, & T. Zumer (2007): "Foreign Direct Investment and Productivity Spillovers: Updated Evidence from Central and Eastern Europe." *Working Paper 2007/08*, Czech National Bank.

Girma, S. & Y. Gong (2008): "FDI, Linkages and the Efficiency of State-Owned Enterprises in China." *Journal of Development Studies* **44(5)**: pp. 728–749.

Girma, S., H. Görg, & M. Pisu (2008): "Exporting, linkages and productivity spillovers from foreign direct investment." *Canadian Journal of Economics* **41(1)**: pp. 320–340.

Girma, S. & K. Wakelin (2007): "Local productivity spillovers from foreign direct investment in the U.K. electronics industry." *Regional Science and Urban Economics* **37(3)**: pp. 399–412.

Goldfarb, R. S. (1995): "The Economist-as-Audience Needs a Methodology of Plausible Inference." *Journal of Economic Methodology* **2(2)**: pp. 201–22.

Gonçalves, J. (2005): *Empresas Estrangeiras e Transbordamentos de Produtividade na Indústria Brasileira: 1997-2000.* Rio de Janeiro: BNDES.

Gorodnichenko, Y., J. Svejnar, & K. Terrell (2007): "When Does FDI Have

Positive Spillovers? Evidence from 17 Emerging Market Economies." *CEPR Discussion Papers 6546*, Centre for Economic Policy Research.

Görg, H. & D. Greenaway (2004): "Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment?" *World Bank Research Observer* **19(2)**: pp. 171–197.

Görg, H. & E. Strobl (2001): "Multinational Companies and Productivity Spillovers: A Meta-analysis." *The Economic Journal* **111(475)**: pp. F723–39.

Hagemejer, J. & M. Kolasa (2008): "Internationalization and economic performance of enterprises: evidence from firm-level data." *National Bank of Poland Working Papers 51*, National Bank of Poland.

Hale, G., C. Long, T. Moran, & H. Miura (2010): "Where to Find Positive Productivity Spillovers from FDI in China: Disaggregated Analysis." *Working papers*, Federal Reserve Bank of San Francisco.

Halpern, L. & B. Muraközy (2007): "Does distance matter in spillover?" *Economics of Transition* **15**: pp. 781–805.

Haskel, J. E., S. C. Pereira, & M. J. Slaughter (2007): "Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?" *The Review of Economics and Statistics* **89(3)**: pp. 482–496.

Havranek, T. (2010): "Rose Effect and the Euro: Is the Magic Gone?" *Review of World Economics* **146(2)**: pp. 241–261.

Havranek, T. & Z. Irsova (2010): "Which Foreigners Are Worth Wooing? A Meta-Analysis of Vertical Spillovers from FDI." *Working Papers 2010/03*, Czech National Bank, Research Department.

Havranek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85(2)**: pp. 234–244.

Havranek, T., Z. Irsova, & K. Janda (2012): "Demand for Gasoline is More Price-Inelastic than Commonly Thought." *Energy Economics* **34(1)**: p. 201–207.

JABBOUR, L. & J. L. MUCCHIELLI (2007): "Technology transfer through vertical linkages: The case of the Spanish manufacturing industry." *Journal of Applied Economics* **10(1)**: pp. 115–136.

JAVORCIK, B. S. (2004): "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages." *American Economic Review* **94(3)**: pp. 605–627.

JAVORCIK, B. S., K. SAGGI, & M. SPATAREANU (2004): "Does it matter where you come from? Vertical spillovers from foreign direct investment and the nationality of investors." *Policy Research Working Paper Series 3449*, The World Bank.

JAVORCIK, B. S. & M. SPATAREANU (2008): "To share or not to share: Does local participation matter for spillovers from foreign direct investment?" *Journal of Development Economics* **85(1-2)**: pp. 194–217.

JORDAAN, J. A. (2008): "Regional foreign participation and externalities: new empirical evidence from Mexican regions." *Environment and Planning A* **40(12)**: pp. 2948–2969.

KELLER, W. (2009): "International Trade, Foreign Direct Investment, and Technology Spillovers." *NBER Working Papers 15442*, National Bureau of Economic Research.

KOLASA, M. (2008): "How does FDI inflow affect productivity of domestic firms? The role of horizontal and vertical spillovers, absorptive capacity and competition." *Journal of International Trade & Economic Development* **17(1)**: pp. 155–173.

KUGLER, M. (2006): "Spillovers from foreign direct investment: Within or between industries?" *Journal of Development Economics* **80(2)**: pp. 444–477.

LE, Q. H. & R. POMFRET (2008): "Technology Spillovers from Foreign Direct Investment in Vietnam: Horizontal or Vertical Spillovers?" *Working Paper 85*, Vietnam Development Forum.

LESHER, M. & S. MIROUDOT (2008): "FDI Spillovers and their Interrelationships with Trade." *OECD Trade Policy Working Papers 80*, OECD, Trade Directorate.

LIANG, F. H. (2008): "Does Foreign Direct Investment Improve the Productivity of Domestic Firms? Technology Spillovers, Industry Linkages, and Firm Capabilities." *Working paper*, Haas School of Business, University of California, Berkeley.

LILEEVA, A. (2006): "Global Links: The Benefits to Domestically-controlled Plants from Inward Direct Investment—The Role of Vertical Linkages." *The Canadian Economy in Transition Research Paper 010*, Statistics Canada, Economic Analysis Division.

LIN, P., Z. LIU, & Y. ZHANG (2009): "Do Chinese domestic firms benefit from FDI inflow? Evidence of horizontal and vertical spillovers." *China Economic Review* **20(4)**: pp. 677–691.

LIU, X., C. WANG, & Y. WEI (2009): "Do local manufacturing firms benefit from transactional linkages with multinational enterprises in China?" *Journal of International Business Studies* **40(7)**: pp. 1113–1130.

LIU, Z. (2008): "Foreign direct investment and technology spillovers: Theory and evidence." *Journal of Development Economics* **85(1-2)**: pp. 176–193.

MANAGI, S. & S. M. BWALYA (2010): "Foreign direct investment and technology spillovers in sub-Saharan Africa." *Applied Economics Letters* **17(6)**: pp. 605–608.

MERLEVEDE, B. & K. SCHOORS (2005): "Conditional Spillovers from FDI Within and Between Sectors: Evidence from Romania." *Working papers*, Örebro University, Sweden.

MERLEVEDE, B. & K. SCHOORS (2007): "FDI and the Consequences: Towards more complete capture of spillover effects." *William Davidson Institute Working Papers Series wp886*, William Davidson Institute at the University of Michigan, Stephen M. Ross Business School.

MERLEVEDE, B. & K. SCHOORS (2009): "Openness, competition, technology and FDI spillovers: Evidence from Romania." *Working Paper 42*, Forum for Research in Empirical International Trade.

MEYER, K. E. & E. SINANI (2009): "When and where does foreign direct investment generate positive spillovers? A meta-analysis." *Journal of International Business Studies* **40(7)**: pp. 1075–1094.

Nelson, J. & P. Kennedy (2009): "The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment." *Environmental & Resource Economics* **42(3)**: pp. 345–377.

Nguyen, A. N., N. Thang, L. D. Trung, N. Q. Pham, C. D. Nguyen, & N. D. Nguyen (2008a): "Foreign Direct Investment in Vietnam: Is There Any Evidence Of Technological Spillover Effects." *Working Papers 18*, Development and Policies Research Center, Vietnam.

Nguyen, C. D., G. Simpson, D. Saal, A. N. Nguyen, & N. Q. Pham (2008b): "FDI Horizontal and Vertical Effects on Local Firm Technical Efficiency." *Working Papers 17*, Development and Policies Research Center, Vietnam.

Qiu, B., S. Yang, P. Xin, & B. Kirkulak (2009): "FDI technology spillover and the productivity growth of China's manufacturing sector." *Frontiers of Economics in China* **4(2)**: pp. 209–227.

Reganati, F. & E. Sica (2007): "Horizontal and Vertical Spillovers from FDI: Evidence from Panel Data for the Italian Manufacturing Sector." *Journal of Business Economics and Management* **8(4)**: pp. 259–266.

Resmini, L. & M. Nicolini (2007): "Productivity Spillovers and Multinational Enterprises: in Search of a Spatial Dimension." *Papers DYNREG10*, Economic and Social Research Institute.

Rose, A. K. & T. D. Stanley (2005): "A Meta-Analysis of the Effect of Common Currencies on International Trade." *Journal of Economic Surveys* **19(3)**: pp. 347–365.

Sasidharan, S. & A. Ramanathan (2007): "Foreign Direct Investment and Spillovers: Evidence from Indian Manufacturing." *International Journal of Trade and Global Markets* **1(1)**: pp. 5–22.

Schoors, K. & B. van der Tol (2002): "Foreign direct investment spillovers within and between sectors: Evidence from Hungarian data." *Working Papers 02/157*, Ghent University, Faculty of Economics and Business Administration, Belgium.

Stancik, J. (2007): "Horizontal and Vertical FDI Spillovers: Recent Evidence from the Czech Republic." *CERGE-EI Working Papers 340*, CERGE-EI, Prague.

Stancik, J. (2009): "FDI Spillovers in the Czech Republic: Takeovers vs. Greenfields." *Economic Papers 369*, The European Comission.

Stanley, T. & H. Doucouliagos (2007): "Identifying and Correcting Publication Selection Bias in the Efficiency-Wage Literature: Heckman Meta-Regression." *Economics Series 2007/11*, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Stanley, T. & H. Doucouliagos (2010): "Picture This: A Simple Graph That Reveals Much Ado About Research." *Journal of Economic Surveys* **24(1)**: pp. 170–191.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

Stanley, T. D., H. Doucouliagos, & S. B. Jarrell (2008): "Meta-regression analysis as the socio-economics of economics research." *The Journal of Socio-Economics* **37(1)**: pp. 276–292.

Sutton, A. J., K. R. Abrams, D. R. Jones, T. A. Sheldon, & F. Song (2000): *Methods for Meta-analysis in Medical Research.* Chichester: Wiley.

Tang, H. (2008): *Essays on international trade and investment.* Ph.D. thesis, Massachusetts Institute of Technology.

Taymaz, E. & K. Yłlmaz (2008): "Foreign Direct Investment and Productivity

Spillovers: Identifying Linkages Through Product-Based Measures." *Working papers*, Koc University.

Tong, S. Y. & Y. Hu (2007): "Productivity Spillovers from FDI: Detrimental or Beneficial? A Study of Chinese Manufacturing." In B. M. Fleisher, H. Li, & S. Song (editors), "Market Development In China: Spillovers, Growth and Inequality," pp. 190–208. Cheltenham, England: Edward Elgar.

Vacek, P. (2007a): "Panel Data Evidence on Productivity Spillovers from Foreign Direct Investment: Firm-level Measures of Backward and Forward Linkages." In "Essays on International Productivity Spillovers," Ph.D. thesis, Cornell University.

Vacek, P. (2007b): "Productivity Spillovers from Foreign Direct Investment: Industry-level Analysis." In "Essays on International Productivity Spillovers," Ph.D. thesis, Cornell University.

Verardi, V. & C. Croux (2009): "Robust regression in Stata." *Stata Journal* **9(3)**: pp. 439–453.

Wang, C. & Z. Zhao (2008): "Horizontal and vertical spillover effects of foreign direct investment in Chinese manufacturing." *Journal of Chinese Economic and Foreign Trade Studies* **1(1)**: pp. 8–20.

Yudaeva, K., K. Konstantin, N. Melentieva, & N. Ponomareva (2003): "Does Foreign Ownership Matter? Russian Experience." *Economics of Transition* **11(3)**: pp. 383–409.

Zajc Kejzar, K. & A. Kumar (2006): "Inward Foreign Direct Investment and Industrial Restructuring: Micro Evidence—The Slovenian Firms' Growth Model." *Proceedings of Rijeka School of Economics: Journal of Economics and Business* **24(2)**: pp. 185–210.

5.A Meta-Analyses for Individual Studies and Countries

Table 5.6: Meta-analyses for individual studies (published papers)

	Backward		Forward		Horizontal		
Study	Eff.	SE	Eff.	SE	Eff.	SE	N
Atallah Murra (2006)	1.281***	0.132	0.848***	0.051	-0.023	0.079	20
Békés *et al.* (2009)	0.030	0.061	0.034**	0.017	0.040***	0.011	9
Blake *et al.* (2009)	0.065	0.040	0.002	0.006	-0.044***	0.012	21
Blalock & Gertler (2008)	0.087***	0.009			-0.009	0.007	10
Blalock & Simon (2009)	0.02	0.014			0.013*	0.007	24
Bwalya (2006)	1.108	0.734			-0.188***	0.067	22
Crespo *et al.* (2009)	0.058	0.149	-0.003	0.060	0.335	0.218	9
Gersl (2008)	1.389	0.926	0.962*	0.569	-0.152	0.203	12
Girma & Wakelin (2007)	0.280***	0.025	0.280***	0.025	0.099***	0.022	45
Girma & Gong (2008)	-0.083	0.112	0.185***	0.050	-0.001	0.003	120
Girma *et al.* (2008)	1.608**	0.712	-3.432	2.724	2.428***	0.736	75
Halpern & Muraközy (2007)	1.464***	0.131	-0.411	0.747	-0.223***	0.053	58
Jabbour & Mucchielli (2007)	0.088*	0.048	0.108***	0.035	-0.058***	0.013	33
Javorcik (2004)	3.267***	0.351	-0.445***	0.132	0.182*	0.096	80
Javorcik & Spatareanu (2008)	0.374***	0.075			-0.234***	0.044	66
Jordaan (2008)	0.625***	0.086	0.625***	0.086	-0.506***	0.061	38
Kolasa (2008)	0.211***	0.049	0.017	0.022	0.040***	0.009	12
Lin *et al.* (2009)	1.373***	0.117	3.553***	0.303	-0.114***	0.037	90
Liu (2008)	-0.174	0.125	0.046	0.094	-0.094*	0.051	18
Liu *et al.* (2009)	0.850***	0.073	1.26***	0.139	-0.010	0.045	108
Managi & Bwalya (2010)	5.086	4.135			7.135***	2.469	6
Qiu *et al.* (2009)	1.761***	0.123	-0.037	0.033	0.682***	0.117	21
Reganati & Sica (2007)	0.073***	0.023			0.079	0.085	6
Resmini & Nicolini (2007)	0.032***	0.005	0.027***	0.005			22
Sasidharan & Ramanathan (2007)	-0.044	0.338			0.050	0.125	6
Wang & Zhao (2008)	4.363***	0.718	4.363***	0.718	0.122***	0.034	14
Yudaeva *et al.* (2003)	-6.111***	1.162	-1.715***	0.256	1.547***	0.252	17
Zajc Kejzar & Kumar (2006)	0.138**	0.057	0.285***	0.060	0.025***	0.006	32

Note: Spillover effects are estimated by the simple random-effects meta-analysis run separately for each study. SE = standard error. N = number of the estimates of spillovers taken from the study.
***, **, and * denote significance at the 1%, 5%, and 10% levels.

Table 5.7: Meta-analyses for individual studies (unpublished papers)

	Backward		Forward		Horizontal		
Study	Effect	SE	Effect	SE	Effect	SE	N
Barrios *et al.* (2009)	0.267	0.173	-0.791***	0.170	0.694***	0.164	71
Blyde *et al.* (2004)	0.375***	0.062	-0.096**	0.042	0.181***	0.057	188
Chang *et al.* (2007)	-0.027***	0.005	0.042***	0.005	0.105***	0.013	112
Damijan *et al.* (2003)	0.092*	0.052	-0.220***	0.083	0.015**	0.006	29
Damijan *et al.* (2008)	0.01	0.027			0.030***	0.011	104
Fernandes & Paunov (2008)			0.125***	0.009			52
Gersl *et al.* (2007)	-0.344	0.471	-1.041**	0.423	-0.065	0.068	153
Gonçalves (2005)	0.668***	0.120					2
Gorodnichenko *et al.* (2007)	0.084***	0.008	0.035***	0.007	0.020***	0.003	243
Hagemejer & Kolasa (2008)	2.919***	0.405	-0.159**	0.071	0.196***	0.032	36
Hale *et al.* (2010)	0.095**	0.041	0.047	0.036			160
Javorcik *et al.* (2004)	4.450***	0.652			0.452***	0.079	24
Le & Pomfret (2008)	1.062***	0.140			-0.825***	0.152	39
Lesher & Miroudot (2008)	-0.341***	0.102	-0.125	0.142	-0.047**	0.023	172
Liang (2008)	-0.216***	0.036	0.438***	0.049	0.008*	0.004	72
Lileeva (2006)	0.126*	0.075	1.544***	0.113	-0.322***	0.037	159
Merlevede & Schoors (2005)	-0.690***	0.167	2.293***	0.457	-0.073	0.166	45
Merlevede & Schoors (2007)	0.097	0.170	0.476***	0.160	-0.044	0.046	60
Merlevede & Schoors (2009)	0.692	1.003	0.181	2.263	2.251***	0.706	42
Nguyen *et al.* (2008a)	-0.158***	0.057	-3.327***	0.293	0.016	0.043	184
Nguyen *et al.* (2008b)	0.097	0.103	-0.487	0.320	-0.024	0.069	20
Schoors & van der Tol (2002)	2.794***	0.244	-3.902***	0.328	0.279***	0.064	54
Stancik (2007)	-1.715***	0.204	-0.279	0.189	-0.158***	0.034	69
Stancik (2009)	-0.787***	0.138	0.322	0.224	-0.023	0.037	84
Tang (2008)	-0.189***	0.043			-0.266***	0.022	257
Taymaz & Yłlmaz (2008)	0.035**	0.015	0.064**	0.029	0.106**	0.052	53
Tong & Hu (2007)	0.228	0.415	0.228	0.415	-0.185	0.325	8
Vacek (2007b)	0.048	0.060	-0.003	0.038	0.013	0.012	48
Vacek (2007a)	0.526***	0.044	-0.001	0.014			92

Note: Spillover effects are estimated by the simple random-effects meta-analysis run separately for each study. SE = standard error. N = number of the estimates of spillovers taken from the study.
***, **, and * denote significance at the 1%, 5%, and 10% levels.

Table 5.8: Meta-analyses for individual countries

	Backward		Forward		Horizontal		
Country	Effect	SE	Effect	SE	Effect	SE	N
Bosnia and Herzegovina	0.553	1.317			-0.268	0.362	8
Bulgaria	-0.333	0.564	-0.501*	0.268	-0.116	0.098	27
Canada	0.126*	0.075	1.544***	0.113	-0.322***	0.037	159
Chile			0.125***	0.009			52
China	0.145***	0.015	0.44***	0.023	-0.004	0.006	1001
Colombia	1.281***	0.132	0.848***	0.051	-0.023	0.079	20
Croatia	0.160	0.108			0.020	0.040	8
Czech Republic	-0.15**	0.063	0.005	0.026	-0.036**	0.014	332
Estonia	0.119	0.253	1.311	1.066	-0.003	0.021	27
Hungary	1.479***	0.121	-0.93***	0.139	-0.023	0.024	148
India	-0.044	0.338			0.050	0.125	6
Indonesia	0.052***	0.011			0.002	0.004	34
Ireland	0.267	0.173	-0.791***	0.170	0.694***	0.164	71
Italy	0.073***	0.023			0.079	0.085	6
Latvia	-0.819*	0.465	0.110	0.579	-0.005	0.023	27
Lithuania	2.845***	0.350	-0.436***	0.129	0.081	0.084	89
Mexico	0.625***	0.086	0.625***	0.086	-0.506***	0.061	57
Poland	1.478***	0.220	-0.092**	0.042	0.099***	0.018	75
Portugal	0.058	0.149	-0.003	0.060	0.335	0.218	9
Romania	0.269**	0.111	1.327***	0.327	0.034	0.055	263
Russian Federation	-6.111***	1.162	-1.715***	0.256	1.547***	0.252	17
Slovakia	0.281*	0.165	-0.442	0.413	0.032	0.027	20
Slovenia	0.127**	0.062	-0.033	0.206	0.011***	0.004	40
Spain	0.088*	0.048	0.108***	0.035	-0.058***	0.013	33
Turkey	0.035**	0.015	0.064**	0.029	0.106**	0.052	53
Ukraine	15.051	12.755			-0.164	0.231	8
United Kingdom	0.293***	0.032	0.279***	0.024	0.104***	0.025	138
Venezuela	0.375***	0.062	-0.096**	0.042	0.181***	0.057	188
Vietnam	0.079	0.049	-3.059***	0.281	-0.038	0.040	243
Zambia	1.108	0.734			-0.188***	0.067	22
Advanced OECD countries[a]	-0.341***	0.102	-0.125	0.142	-0.047**	0.023	172
Transition countries[b]	0.085***	0.008	0.035***	0.007	0.02***	0.003	231

Note: Spillover effects are estimated by the simple random-effects meta-analysis run separately for each country. Meta-analyses for countries for which we have less than five estimates are not reported, but are available on request. SE = standard error. N = number of the estimates of spillovers for the country.
***, **, and * denote significance at the 1%, 5%, and 10% levels.

[a] Austria, Belgium, Finland, France, Germany, Luxembourg, Netherlands, Norway, Sweden.
[b] Albania, Georgia, Kazakhstan, Serbia.

Chapter 6

Cross-Country Heterogeneity in Intertemporal Substitution

Abstract

We collect 2,735 estimates of the elasticity of intertemporal substitution in consumption from 169 published studies that cover 104 countries during different time periods. The estimates vary substantially from country to country, even after controlling for 30 aspects of study design. Our results suggest that income and asset market participation are the most effective factors in explaining the heterogeneity: households in rich countries and countries with high stock market participation substitute a larger fraction of consumption intertemporally in response to changes in expected asset returns. Micro-level studies that focus on sub-samples of rich households or asset holders also find systematically larger values of the elasticity.

Keywords: Elasticity of intertemporal substitution, consumption, meta-analysis, Bayesian model averaging

JEL Codes: C83, D91, E21

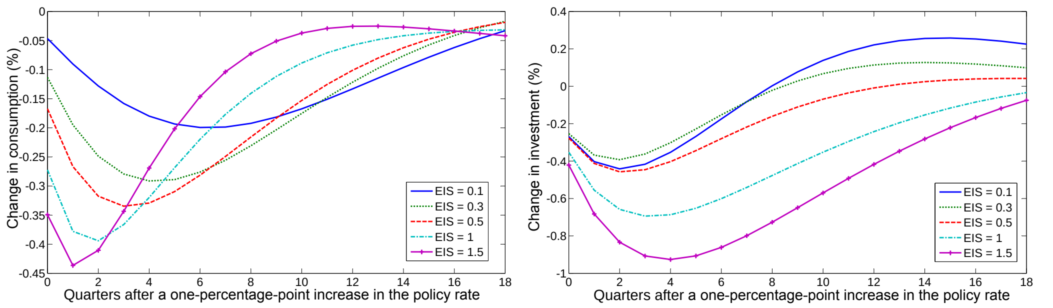
This paper is a joint work with Tomas Havranek, Roman Horvath, and Marek Rusnak. An online appendix with data, code, and a list of studies included in the meta-analysis is available at `meta-analysis.cz/substitution`. Havranek, Irsova, and Rusnak acknowledge support from the Grant Agency of Charles University in Prague (grant #554213). We thank Jan Babecky, Michal Bauer, Iftekhar Hasan, Jiri Schwarz, and seminar participants at CERGE-EI, Charles University, and the Czech National Bank for their helpful comments. The paper is currently at the revise and resubmit stage at the Journal of International Economics.

6.1 Introduction

The elasticity of intertemporal substitution in consumption (EIS) reflects households' willingness to substitute consumption between time periods in response to changes in the expected real interest rate. Therefore it represents a crucial parameter for a wide range of economic models involving intertemporal choice, from modeling the behavior of aggregate savings and the impact of fiscal policy to computing the social cost of carbon emissions, and has been estimated by hundreds of researchers. Figure 6.1 illustrates how the elasticity matters for the modeled effects of monetary policy: we use the popular model of Smets & Wouters (2007), vary the calibrated value of the EIS, and for different values of the EIS plot the impulse responses of consumption and investment to a one-percentage-point monetary policy shock. It is apparent that the modeled development of these aggregates depends strongly on the value of the elasticity of intertemporal substitution.

Figure 6.1: The elasticity of intertemporal substitution matters

(a) Consumption (b) Investment



Notes: The figure shows simulated impulse responses to a one-percentage-point increase in the monetary policy rate. We use the popular model developed by Smets & Wouters (2007) and vary the value of the elasticity of intertemporal substitution while leaving all other parameters calibrated at the posterior values from Smets & Wouters (2007). For the simulations we use Matlab code from The Macroeconomic Model Data Base (Wieland *et al.* 2012).

The figure shows impulse responses for the EIS calibrated between 0.1 and 1.5, and in the literature we indeed encounter such large differences in calibrations of the elasticity. The most cited empirical study estimating the elasticity, Hall (1988), who concludes that the EIS is not likely to be larger than 0.1, has influenced many researchers. Some studies use a value of 0.2 (Chari *et al.* 2002; House & Shapiro 2006;

Piazzesi *et al.* 2007), or a value of 0.5 (Jin 2012; Trabandt & Uhlig 2011; Rudebusch & Swanson 2012), or a value of 2 (Ai 2010; Barro 2009; Colacito & Croce 2011), to name but a few recent examples of different calibrations. The reason for the different calibrations is differences in the results of empirical studies on the EIS. For example, the standard deviation of the estimates reported by the 33 studies in our sample which were published in the top five general interest journals is 1.4, outliers excluded. Most commentators would agree with Ai (2010, p. 1357), who starts his discussion of calibration by noting that "empirical evidence on the magnitude of the EIS parameter is mixed."

In this paper we collect 2,735 estimates of the elasticity of intertemporal substitution reported in 169 studies and review the literature quantitatively using meta-analysis methods. Meta-analysis, which has been employed in economics by Card & Krueger (1995), Ashenfelter *et al.* (1999), Stanley (2001), Disdier & Head (2008), and Chetty *et al.* (2011), among others, allows us to examine systematically the influence of methodology on the results. In this framework we can address the challenge put forward by an early survey of the empirical evidence from consumption Euler equations (Browning & Lusardi 1996, p. 1833): "It is frustrating in the extreme that we have very little idea of what gives rise to the different findings. (...) We still await a study which traces all of the sources of differences in conclusions to sample period; sample selection; functional form; variable definition; demographic controls; econometric technique; stochastic specification; instrument definition; etc."

While controlling for differences in methodology, we focus on explaining country-level heterogeneity. The studies in our sample provide us with estimates of the EIS for 104 countries, and we show that the mean values reported for the countries vary substantially. We build on the literature that explores the heterogeneity in the EIS at the micro level. For example, Blundell *et al.* (1994) and Attanasio & Browning (1995) suggest that rich households tend to show a larger elasticity of intertemporal substitution, and we examine whether GDP per capita is associated with the mean EIS reported for the country. Mankiw & Zeldes (1991) and Vissing-Jorgensen (2002) find a larger elasticity for stockholders than for non-stockholders, and we explore the relationship between stock market participation and the elasticity of intertemporal substitution at the country level. Bayoumi (1993) and Wirjanto (1995), among others, indicate that liquidity-constrained households show a smaller EIS, and we examine whether ease of access to credit helps explain the cross-country variation in the elasticity. More details on factors potentially causing heterogeneity in the EIS are available in Section 6.3.

The mean estimate of the elasticity of intertemporal substitution reported in empirical studies is 0.5, but we show that cross-country differences are important. Since it is often unclear which aspects of methodology should matter for the magnitude of the estimated EIS, we include all 30 that we collect and employ Bayesian model averaging (Raftery *et al.* 1997) to deal with the resulting model uncertainty. Our findings suggest that a larger EIS is associated with higher per capita income of the country, and especially with higher stock market participation. According to our baseline model, a 10-percentage-point increase in the rate of stock market participation is associated with an increase in the EIS of 0.24. Moreover, wealth and asset market participation are also important at the micro level: studies estimating the EIS using a sub-sample of rich households or asset holders find on average an EIS larger by 0.21.

The remainder of the paper is structured as follows. Section 6.2 explains how we collect data from studies estimating the elasticity. Section 6.3 discusses the reasons for including variables that may explain the differences in the reported estimates of the EIS. Section 6.4 describes the results, while Section 6.5 provides robustness checks. Section 6.A lists mean values of the EIS reported for various countries and summary statistics of all variables used in our analysis. Section 6.B provides diagnostics on Bayesian model averaging. An online appendix with data, code, and a list of studies included in the meta-analysis is available at `meta-analysis.cz/substitution`.

6.2 Estimates of the Elasticity

To estimate the EIS, researchers often follow Hall (1988) and use the log-linearized consumption Euler equation. That is, they regress consumption growth on the intertemporal price of consumption, the real rate of return:

$$\Delta c_{t+1} = \alpha_i + EIS \cdot r_{i,t+1} + \epsilon_{i,t+1}. \tag{6.1}$$

Here Δc_{t+1} denotes consumption growth at time $t + 1$, $r_{i,t+1}$ denotes the real return on asset i at time $t + 1$ (for instance the stock market return or treasury bill return), and $\epsilon_{i,t+1}$ denotes the error term. The error term is correlated with $r_{i,t+1}$, and researchers thus use instruments for $r_{i,t+1}$, typically including the values of asset returns and consumption growth known at time t. There are of course many potential modifications to (6.1), many ways in which it can be estimated, and many different data that can be used in the estimation; we discuss these issues in detail in Section 6.3 and control for the context in which researchers obtain their estimates.

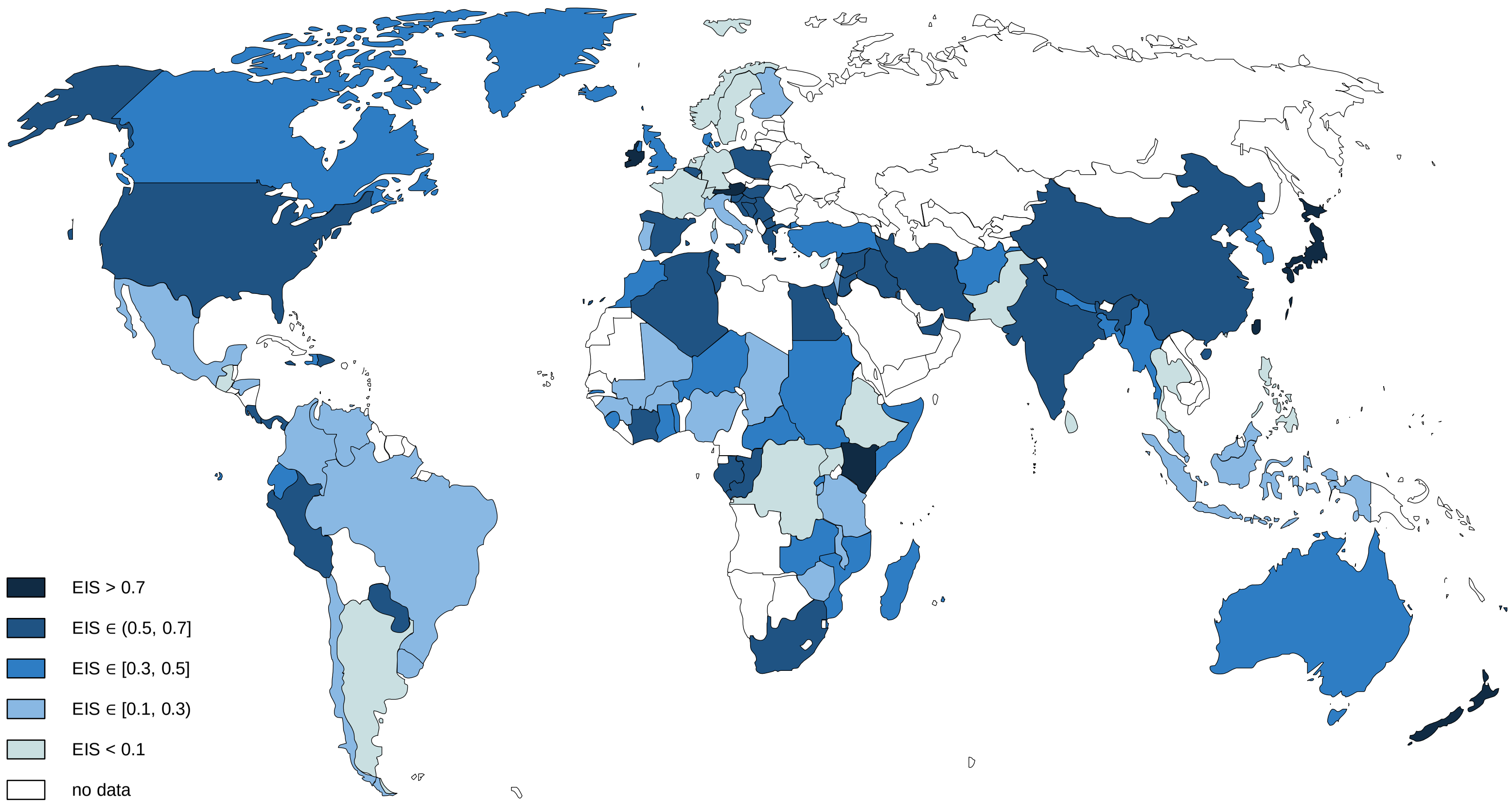
The first and crucial step of meta-analysis is the selection of studies that are included. We start with an extensive search in Google Scholar (the search query and the list of studies are available in the online appendix). There are thousands of papers on the topic, so a good search query is needed to identify studies that are likely to contain empirical estimates of the EIS. We adjust our query until it includes most of the well-known empirical papers among the top 50 hits. For the selection of studies we prefer Google Scholar to other databases commonly used in meta-analysis, such as EconLit or Scopus, because Google Scholar provides powerful fulltext search.

The search yields about 1,500 hits in total, but on closer examination we find that papers identified in the bottom half of the search list are unlikely to contain usable empirical estimates of the EIS. We read the abstracts of the first 700 papers to see which can be included in the meta-analysis, and it seems that more than 300 studies contain usable estimates of the EIS. At this point it is clear that to capture the context in which researchers obtain the estimates we have to collect about 30 variables reflecting methodology. Since a typical study (especially a typical working paper) reports many different estimates (using different sets of instrumental variables, for example), we find it unfeasible to include all studies and decide to focus on published studies only and read these studies in detail. An alternative solution is to select just one representative estimate from each study, published or unpublished, and discard the other estimates, but often it is unclear what the preferred estimate would be. We stop the search on January 1, 2013 and identify 169 published studies that provide estimates of the EIS and detailed information on methodology.

Aside from saving us several months of work, the restriction of the sample to published studies has two additional benefits. First, publication status is a simple indicator of quality because published studies are peer-reviewed. Second, published papers are typically better written and typeset, which makes the collection of data easier and reduces the danger of mistakes. But even when we focus solely on published papers, we have to collect about 80,000 data points by hand (the published literature provides 2,735 estimates of the EIS and for each we collect 30 aspects of methodology). Two of the co-authors, therefore, collect the data simultaneously and check the resulting data set for errors. The final database used in the paper is available in the online appendix. Judging from the surveys of meta-analyses by Nelson & Kennedy (2009) and Doucouliagos & Stanley (2013) we believe this paper is the largest meta-analysis conducted in economics so far.

Out of the 169 studies included in the meta-analysis, 33 are published in the top five journals in economics, which underlines the importance of the EIS and the amount of research dedicated to its estimation. All studies combined receive on average more than two thousand citations per year in Google Scholar, which indicates that the estimates are heavily used. Our sample includes studies published over three decades: from 1981 to 2012; the median study uses data from 1970 to 1994 and provides 8 estimates of the elasticity. The estimates span 104 different countries, even though about half of all estimates are computed for the US. The mean reported estimate of the EIS is 0.5—for this and all other computations we exclude estimates that are larger than 10 in absolute value (2.5% of the data). Such large estimates seem implausible, but the threshold is arbitrary. In Section 6.5 we explain that the

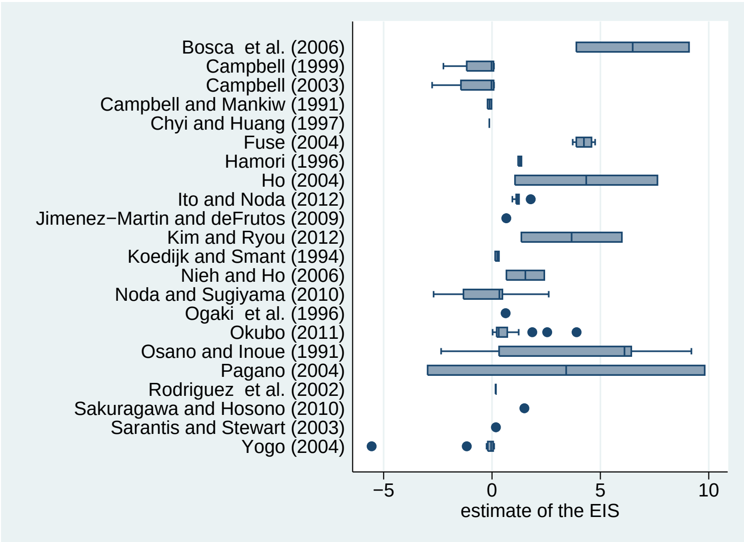
Figure 6.2: Country heterogeneity in the EIS



Notes: For each country the figure depicts the mean estimate of the EIS reported in the literature; numerical values are provided in Table 6.7. Estimates larger than 10 in absolute value are excluded.

choice of threshold does not affect our results much. Finally, when each study is given the same weight (as opposed to each estimate being given the same weight), the mean EIS is 0.7. This is close to, for example, the baseline calibration of 2/3 used by Smets & Wouters (2007).

Figure 6.3: Method heterogeneity in the EIS for Japan



Notes: The figure is a box plot of estimates of the EIS corresponding to Japan that are reported in the studies in our sample. Estimates larger than 10 in absolute value are excluded.

But the worldwide mean represents a poor guide for the calibration of the EIS in most countries, as Figure 6.2 illustrates (numerical values for the countries are provided in Table 6.7 in the Appendix). The estimated EIS differs a lot across countries, typically lying between 0 and 1. Such heterogeneity can make a big difference to the modeled effectiveness of monetary policy, among other things, as we showed in Figure 6.1. For some countries only a handful of estimates are available, so some of the country averages we report may be quite imprecise and influenced by the estimation method. Nevertheless, for six countries we have more than 50 estimates (the least covered of these countries is Sweden, with 63 estimates reported in 11 studies). Among these countries we find the largest EIS for Japan (0.9), followed by the US (0.6), the UK (0.5), Canada (0.4), Israel (0.2), and Sweden (0.1). The cross-country heterogeneity in the estimated EIS is substantial and calls for an explanation.

When looking for the sources of cross-country heterogeneity, however, it is also important to take into account that researchers employ different methods to estimate the EIS. Figure 6.3 shows how the reported EIS differs across studies even if it is estimated for the same country. For illustration we select Japan, which is the third most often examined country in the literature (after the US and the UK). Dozens of studies estimate the elasticity for the US and the UK and it would be difficult to squeeze them into a box plot, but the conclusion would be the same even for these countries. We see that individual studies report very different estimates and often the within-study distributions of the estimates do not overlap. Therefore, in all the estimations we also control for the methodology employed by the researchers.

6.3 Why Do the Estimates Differ?

We consider five country characteristics that may influence the reported magnitude of the EIS:

Income Most studies examining heterogeneity in the EIS focus on the role of income. The hypothesis states that poor consumers substitute less consumption intertemporally because their consumption bundle contains a larger share of necessities, which are more difficult to substitute between time periods compared with luxury goods. Moreover, if subsistence requirements represent an important portion of the poor's consumption, the poor have limited discretion for intertemporal substitution in consumption. This hypothesis has been supported by analyses of micro data (for example, Blundell *et al.* 1994; Attanasio & Browning 1995), as well as cross-country data (Atkeson & Ogaki 1996; Ogaki *et al.* 1996). We use GDP per capita to capture the differences in income across countries.

Asset market participation We expect households participating in asset markets to be more willing to substitute consumption intertemporally. Exposure to the stock market, for example, may be correlated with households' awareness of the payoffs from intertemporal substitution and, in general, with the forward-looking nature of their consumption. Moreover, Attanasio *et al.* (2002) and Vissing-Jorgensen (2002) argue that consumption Euler equations are not valid for households not participating in the corresponding asset market, and find larger estimates of the EIS for stockholders and bondholders compared with households that do not own these assets. Similarly, Mankiw & Zeldes (1991) find a larger EIS for stockholders than for other households. To capture this country characteristic we use the database of stock market participation developed by Giannetti & Koskinen (2010).

Liquidity constraints Liquidity-constrained households have less opportunities for intertemporal substitution in consumption (Wirjanto 1995). The resulting consumption of liquidity-constrained households may be linked to income, as it is for the rule-of-thumb consumers of Campbell & Mankiw (1989), and lacks the forward-looking element of the response to the expected real rate of return. Bayoumi (1993), for example, finds that financial deregulation in the UK brought a substantial increase in the proportion of households with a positive EIS. Attanasio (1995) provides a survey of the literature on the effects of liquidity constraints on intertemporal consumption choice. To capture liquidity constraints we use two alternative measures: credit availability defined as the ease of access to loans and reported by the Global Competitiveness Report, and a measure of financial reform reported by the IMF (Abiad *et al.* 2010).

Asset return Almost all estimations and applications of the EIS assume the elasticity to be constant with respect to the rate of return of the asset in question. In a recent paper, however, Crossley & Low (2011) reject the hypothesis of a constant EIS. To see whether the estimated EIS differs systematically for countries with different returns, we include a measure of the real interest rate defined as the lending rate adjusted for inflation as measured by the GDP deflator.

Culture and institutions The willingness of households to substitute consumption into an uncertain future may be associated with culture and institutions. For example, Porta *et al.* (1998) suggest that institutions have an important influence on financial decisions. It has also been found that trust, or social capital more generally, is an important factor for stock market participation and financial development (Guiso *et al.* 2004; 2008). Moreover, a large cross-country survey on time discounting and risk preferences (Wang *et al.* 2011; Rieger *et al.* 2011) shows the importance of cultural differences. To capture the economic culture of the country we use two measures: the rule of law index (taken from the World Bank Global Governance Indicators), which captures the extent to which people have confidence in the rules of society, and the index of generalized trust in society (Bjoernskov & Meon 2013).

A detailed description and summary statistics for each variable used in our analysis are reported in Table 6.8 in the Appendix. A few difficult issues of data collection are worth discussing at this point. First, some variables are not available for all 104 countries in our data set. Data on stock market participation are available for only 28 countries, which we call "core countries" in the analysis, and we also conduct a separate set of regressions without the variable on stock market participation (and, therefore, using almost all countries in the data set). Second, a few estimates of the EIS use data from several countries; for example, the euro area. We keep such estimates in the data set and compute average values of the corresponding country-level characteristics. Third, different studies use data from different time periods to estimate the EIS. Whenever possible, we compute the average of the country characteristic corresponding to the data period. For example, if a study uses data from 1980 to 1994, we use the average value of the real interest rate of that period. This adjustment significantly increases the variation in country-level variables.

We also consider 30 variables reflecting the different aspects of methodology used to estimate the EIS. For ease of exposition we divide these method choices into variables reflecting the definition of the utility function (5 aspects), data characteristics (6 aspects), general design of the analysis (7 aspects), the definition of main variables (4 aspects), estimation characteristics (4 aspects), and publication characteristics (4 aspects).

Utility function An important feature of studies estimating the EIS is whether the elasticity is separated from the coefficient of relative risk aversion. Only about 5% of all the estimates in our sample estimate the parameters separately, usually employing the utility function put forward by Epstein & Zin (1989). Habits in consumption are assumed by 4% of researchers. Some studies assume non-separability between durables and non-durables (4% of estimates), following Ogaki & Reinhart (1998), who argue that assuming separability can produce a downward bias in the estimate of the elasticity. A similar fraction of studies allow for non-separability between private and public consumption, while 5% of studies allow for non-separability between tradable and non-tradable goods.

Data The studies differ greatly in the number of cross-sectional units (usually households or countries) used in the estimation and in the length of the time span of the data. We also include a variable reflecting the average year of the data period to see whether there is a trend in the estimated EIS over time. We include a dummy variable for studies using micro data (about 20% of our data set). Many authors (for example, Attanasio & Weber 1993) argue that estimating Euler equations on macro data can lead to biased results because of the omission of demographic factors. Moreover, we include dummy variables reflecting the frequency of the data used for the estimation. Most studies use quarterly data (57%); some employ monthly data (10%). Annual data are typically used by micro studies.

Design We include a dummy variable for studies using synthetic cohort data (about 5% of our data set). Most authors assume a time-additive utility function, which results in the EIS being equal to the inverse of the coefficient of relative risk aversion. Some studies focusing on risk preferences regress asset returns on consumption growth and report the inverse of the EIS (almost a third of all the studies in our data set). Nevertheless, Campbell (1999) notes that using the asset return as the response variable may aggravate the problem of weak instruments in estimating the parameter.

To see whether this method choice has a systematic effect on the results, we include a dummy variable called *Inverse estimation*.

As we noted earlier, some micro studies on the EIS explore potential heterogeneity in the parameter; they typically estimate the elasticity for different subsets of households. The definition of subsets differs, but researchers usually ask whether richer households or households participating in asset markets show a larger elasticity of intertemporal substitution. To capture this effect we include a dummy variable *Asset holders*. Next, Campbell & Mankiw (1989), among others, show that because of the time aggregation of consumption the instrument set for asset returns should not contain first lags of variables. But still about 30% of all the estimates are computed using first lags of variables among the instruments.

Gruber (2006) stresses that studies using micro data should include year fixed effects for the identification to come from cross-sectional variation and not from time series variation correlated with consumption. Nevertheless, 3% of the studies in our data set use data from the Panel Study of Income Dynamics but do not include year fixed effects. About a quarter of the studies include income in the estimation to test for excess sensitivity of consumption to current income, and we control for this aspect of methodology as well. We also include the number of demographic controls used in micro studies to explain household-level variation in consumption.

Variable definition Most studies use non-durable consumption as the response variable, but some 20% of the estimates are computed using total consumption. About 6% of studies use food as a proxy for consumption, which according to Attanasio & Weber (1995) can produce biased estimates if food is not separable from other types of consumption. The asset return is typically defined as the interest rate on treasury bills, but almost 20% of studies use the stock market return. Mulligan (2002), however, explains that the rate of return should be measured as the return on a representative unit of capital, and we include a dummy variable for this aspect of methodology.

Estimation We have noted that the log-linearized consumption Euler equation is the favorite framework for estimation of the EIS. But Carroll (2001), for example, criticizes the common practice on the grounds that higher-order terms may be endogenous to omitted variables in the regression resulting from the log-linear Euler equation. Thus we include a dummy variable for studies using the exact Euler equation to see whether log-linearization affects the estimates of the elasticity in a systematic way. Next, the regression parameters are typically estimated using GMM, but a third of studies use two-stage least squares, and 10% of studies disregard endogeneity and employ OLS.

Publication characteristics Some novel methods are employed by only a few studies and their influence on the results cannot be examined in a meaningful way using meta-analysis. For this reason we also include variables reflecting the quality of studies not captured by the method variables introduced above. We include publication year to capture innovations in methodology, the number of citations of the study in Google Scholar, the recursive RePEc impact factor of the journal, and a dummy variable for studies published in the top five general interest journals in economics. The data on citations and impact factors were collected on January 31, 2013.

6.4 Meta-Regression Analysis

Our intention is to explore whether the country characteristics described in the previous section are associated with the reported EIS, but also to control for the type of methodology used in the studies. That is, we employ the following "meta-regression":

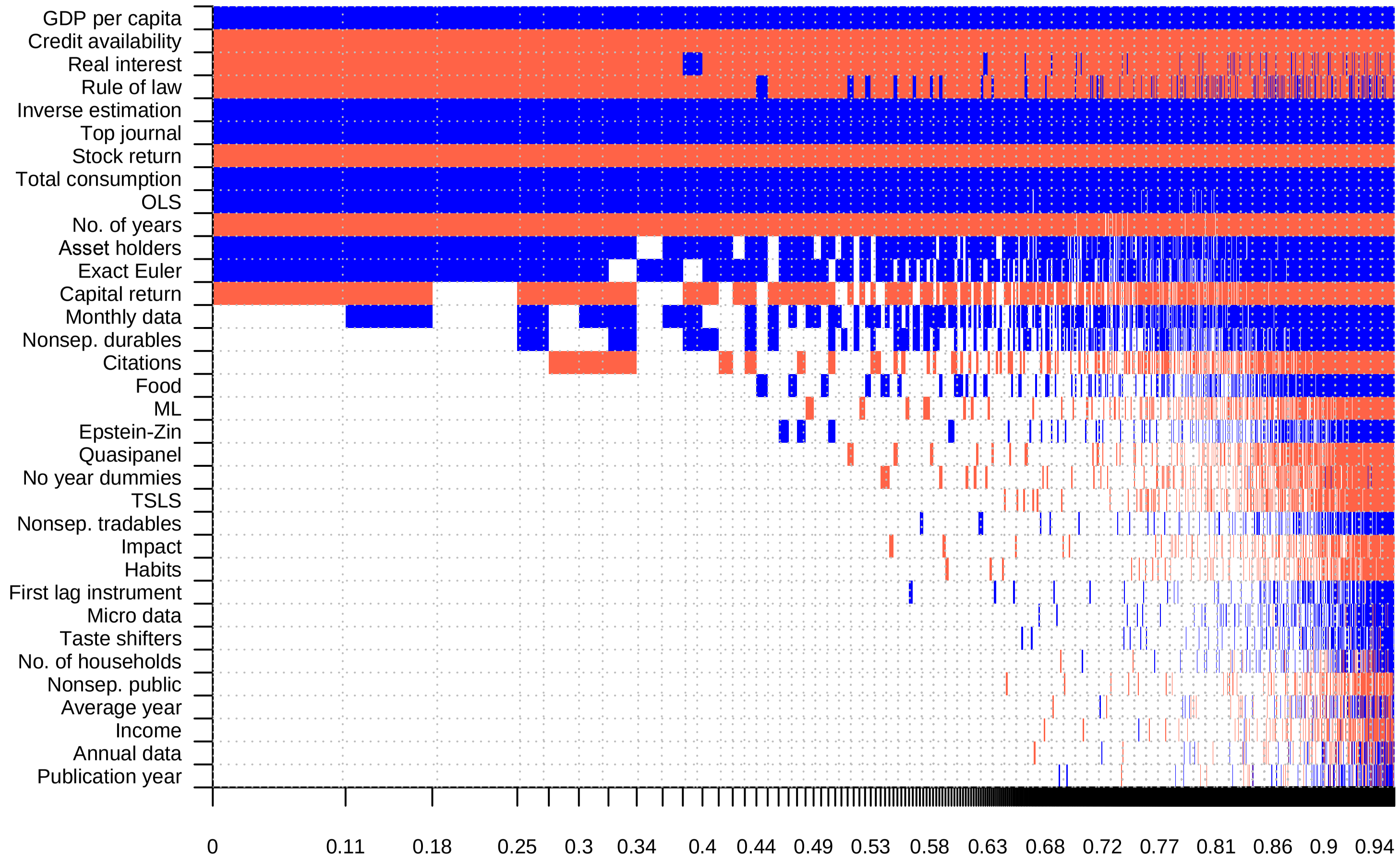
$$EIS_k = a + \boldsymbol{\beta} \cdot \mathit{Country\ variables}_k + \boldsymbol{\gamma} \cdot \mathit{Method\ variables}_k + \theta_k. \tag{6.2}$$

The problem is that there are 30 method variables and it is not clear which ones should be included. We cannot include all of them in an OLS regression because the specification would contain many redundant variables. Some meta-analysts use sequential t-tests to exclude the least significant variables, but such an approach is not statistically valid. In this paper we opt for a technique designed to tackle such regression model uncertainty: Bayesian model averaging (BMA). BMA runs many regressions with different subsets of the explanatory variables on the right-hand side and then constructs a weighted average over these regressions (aside from a robustness check, we always include the country-level variables in all BMA regressions). For applications of BMA in economics, see, for instance, Fernandez *et al.* (2001); Ciccone & Jarocinski (2010); Moral-Benito (2012). Because model uncertainty is inevitable in meta-analysis (it is usually unclear whether some aspects of methodology could influence the results in a systematic way, and the potential aspects are many), BMA has also been frequently used in this field (Moeltner & Woodward 2009; Irsova & Havranek 2013; Havranek & Rusnak 2013).

Bayesian model averaging is described in detail by Feldkircher & Zeugner (2009), for instance, and here we only give intuition for the technical terms needed for the evaluation of the results. The weights used in the BMA estimation are called *posterior model probabilities* and capture how well individual regressions fit the data—thus the weights are analogous to adjusted R-squared or information criteria used in frequentist econometrics. For each variable the sum of the posterior probabilities of models in which the variable is included indicates the so-called *posterior inclusion probability*, which is analogous to statistical significance. If the posterior inclusion probability of a variable is close to one, almost all models that are effective in explaining the variance in the reported EIS include that variable. BMA provides us with a large number of regressions, and from these we can compute for each variable the *posterior coefficient distribution*. The posterior coefficient distribution gives us the posterior mean (analogous to the estimate of a regression coefficient) and posterior standard deviation (analogous to the standard error of an estimated regression parameter).

Because we have 30 method variables, there are 2^{30} potential regressions with different combinations of the method variables. To compute all these regressions would take several weeks, so we opt for the Metropolis-Hasting algorithm, a Markov chain Monte Carlo method. The Metropolis-Hastings algorithm walks through the most

Figure 6.4: Model inclusion, all countries



Notes: Response variable: estimate of the elasticity of intertemporal substitution. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures cumulative posterior model probabilities. Only the 5,000 models with the highest posterior model probabilities are shown.

important part of the model mass—the models with high posterior model probabilities. For all BMA estimations we use one million burn-ins and two million iterations to ensure a good degree of convergence. We employ the beta-binomial prior advocated by Ley & Steel (2009): the prior model probabilities are the same for all possible model sizes. We set the Zellner's g prior following Fernandez *et al.* (2001). These priors are quite conservative and reflect the fact that we know little about the true model size and parameter signs. In the next section, however, we check if our results are robust to a different choice of priors. All of the computations are performed using the R package `bms` available at `bms.zeugner.eu`. Codes for all our estimations are available in the online appendix.

In our first BMA estimation we do not include *stock market participation*, which is available for only 28 countries, and use data for as many countries as possible. The estimation is illustrated in Figure 6.4. The columns in the figure denote individual models; the variables are sorted by posterior inclusion probability in descending order. A blue cell (darker in grayscale) implies that the variable is included and its estimated sign is positive. A red color (lighter in grayscale) implies that the variable is included and the estimated sign is negative. Blank cells imply that the corresponding variable is not included in the model. Only the $5,000$ models with the highest posterior model probabilities are shown, but we can see that they capture almost all of the cumulative model probabilities.

The best models in terms of posterior probabilities are depicted on the left. The very best one includes only 9 out of the 30 method variables at our disposal; the variables included are *inverse estimation*, *top journal*, *stock return*, *total consumption*, *OLS*, *no. of years*, *asset holders*, *exact Euler*, and *capital return*. *Monthly data* is not included in the best model, but it belongs to most of the other good models, and has a posterior inclusion probability larger than 0.5. All other method variables have posterior inclusion probabilities below 0.5, which indicates that they do not matter much for the magnitude of the estimated elasticity. Concerning the country-level variables (which are included in all models), we can see that *GDP per capita* and *credit availability* have the same estimated influence on the EIS no matter what

Table 6.1: Explaining the differences in the estimates of the EIS, all countries

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of the EIS	Post. mean	Post. std. dev.	PIP	Coef.	Std. er.	p-value
Country characteristics						
GDP per capita	0.134	0.074	1.000	0.126	0.084	0.138
Credit availability	-0.037	0.059	1.000	-0.033	0.055	0.553
Real interest	-0.005	0.007	1.000	-0.003	0.006	0.635
Rule of law	-0.020	0.092	1.000	-0.019	0.074	0.800
Utility						
Epstein-Zin	0.018	0.074	0.069			
Habits	-0.004	0.032	0.021			
Nonsep. durables	0.122	0.199	0.309			
Nonsep. public	-0.001	0.019	0.012			
Nonsep. tradables	0.006	0.043	0.027			
Data						
No. of households	0.000	0.003	0.012			
No. of years	-0.201	0.055	0.982	-0.196	0.048	0.000
Average year	0.015	0.940	0.012			
Micro data	0.002	0.026	0.017			
Annual data	0.000	0.008	0.010			
Monthly data	0.160	0.167	0.531	0.263	0.090	0.004
Design						
Quasipanel	-0.015	0.068	0.059			
Inverse estimation	0.530	0.067	1.000	0.512	0.137	0.000
Asset holders	0.349	0.181	0.849	0.421	0.089	0.000
First lag instrument	0.002	0.015	0.021			
No year dummies	-0.027	0.131	0.054			
Income	0.000	0.008	0.011			
Taste shifters	0.001	0.011	0.015			
Variable definition						
Total consumption	0.373	0.085	0.997	0.379	0.102	0.000
Food	0.051	0.147	0.141			
Stock return	-0.344	0.077	0.999	-0.385	0.163	0.021
Capital return	-0.207	0.148	0.723	-0.288	0.077	0.000
Estimation						
Exact Euler	0.219	0.131	0.792	0.283	0.244	0.250
ML	-0.023	0.084	0.085			
TSLS	-0.006	0.035	0.043			
OLS	0.420	0.111	0.984	0.440	0.119	0.000
Publication						
Publication year	0.018	0.843	0.010			
Citations	-0.018	0.032	0.268			
Top journal	0.482	0.085	1.000	0.442	0.074	0.000
Impact	-0.001	0.005	0.025			
Constant	-0.579	NA	1.000	-0.330	0.874	0.706
Observations	2,526			2,526		

Notes: EIS = elasticity of intertemporal substitution. PIP = posterior inclusion probability. Country characteristics are always included in all models of the BMA. In the frequentist check we only include method characteristics with PIP > 0.5. Standard errors in the frequentist check are clustered at the country level. More details on the BMA estimation are available in Table 6.9 and Figure 6.7.

method variables are included. In contrast, the estimated signs for *real interest* and *rule of law* are unstable and depend on the specification of the model.

The numerical results of the BMA estimation are summarized in Table 6.1. For each variable we report the estimated posterior mean for the regression parameter and the corresponding posterior standard deviation together with the posterior inclusion probability (for country-level variables the posterior inclusion probability is one by definition). In the right-hand part of the table we report the results of the frequentist check of our BMA estimation; that is, we also run a simple OLS. In the OLS we only include variables that proved to be relatively important in the BMA exercise (those with posterior inclusion probabilities above 0.5) and cluster the standard errors at the country level. We can see that the results of the frequentist check are very similar to the BMA results. Diagnostics of the BMA estimation are available in Table 6.9 and Figure 6.7 in the Appendix.

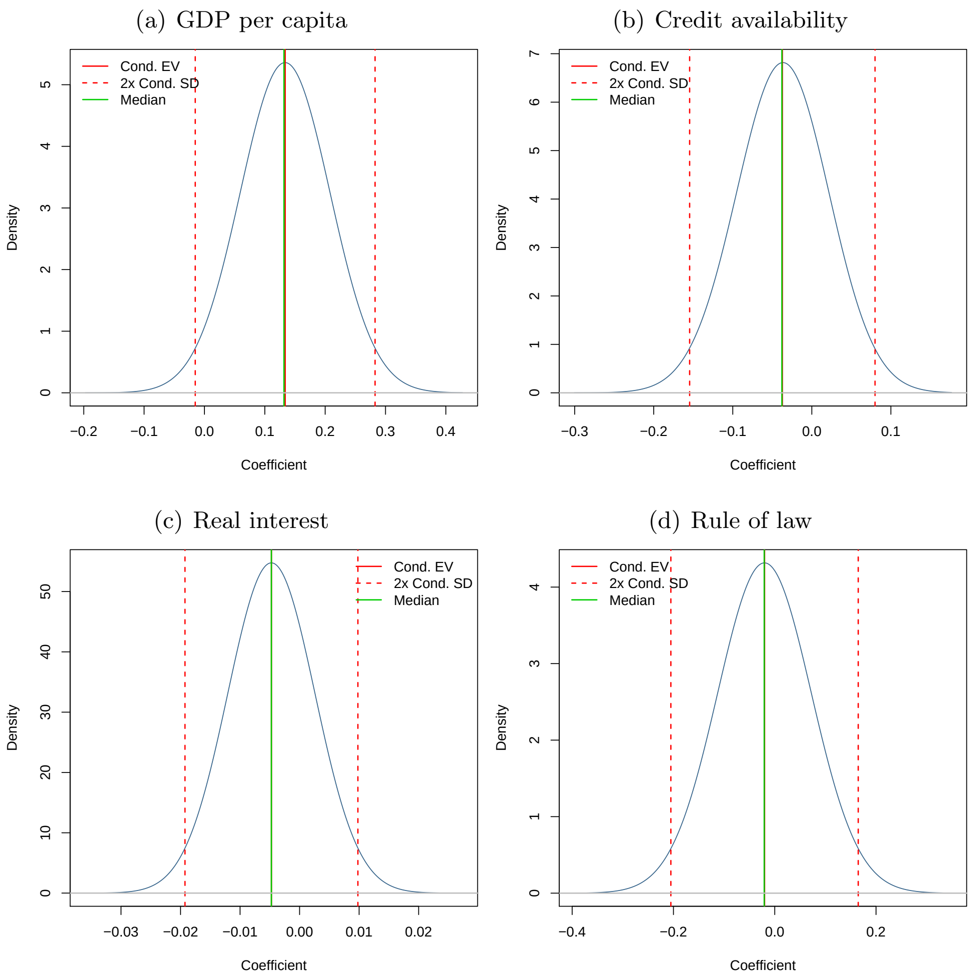
Concerning method variables, our results suggest that the type of utility function does not affect the reported estimates of the EIS in a systematic way. On the other hand, we find that certain aspects of the data are important, namely, that studies using longer time series report smaller estimates of the elasticity and that monthly frequency of data is associated with larger estimates. Both these effects, however, are rather small. An important aspect of study design is whether the EIS is estimated directly in a regression with consumption growth as the response variable or if the inverse of the EIS is estimated in a regression where asset return is on the left-hand side. In the latter case the implied elasticity tends to be larger on average by 0.5, which is a significant difference considering that the mean of all the reported estimates is 0.5 and the practical relevance of such changes of the EIS is large, as illustrated in Figure 6.1.

When the elasticity of intertemporal substitution is estimated for a sub-sample of rich households or stockholders, the estimate tends to be substantially larger as well: by 0.35. Thus poor households and non-asset holders seem to display a significantly smaller EIS, which is in line with Mankiw & Zeldes (1991), Blundell *et al.* (1994), and Vissing-Jorgensen (2002), among others. The definitions of the two main variables in the consumption Euler equations—consumption and asset return—are important as well. When total consumption is used instead of non-durable consumption, the study is likely to find a larger EIS. Also, the use of bond returns as the measure of asset returns, in contrast to the use of stock returns or returns on a unit of capital, is associated with a larger reported EIS.

Studies that estimate the exact consumption Euler equation (that is, studies that do not use log-linear approximation) usually report a larger elasticity. Failure to acknowledge endogeneity when regressing consumption growth on asset returns results in substantial overestimation of the EIS: by about 0.4. Finally, our results also indicate that studies published in the top five general interest journals in economics tend to report estimates of the EIS larger by 0.5 compared with studies published in other journals. The difference may reflect aspects of quality that are not captured by the other variables we collected. Papers published in top journals often present novel methodology, and method aspects that have only been used by a few studies are difficult to examine in a meta-analysis framework.

The country-level variables, which are the main focus of our paper, are included in all the regressions, so for these variables the posterior inclusion probabilities reported in Table 6.1 are not informative. Instead we need to look at the posterior distribution of the regression coefficients reported in Figure 6.5. From the figure we can see that the estimated regression parameters for *credit availability*, *real interest*, and *rule of law* are close to zero. The dashed lines denote values that lie two standard deviations from the mean of the estimated regression parameter; therefore, they can be interpreted as analogous to 95% confidence intervals in frequentist econometrics. Even for *GDP per capita* the interval includes zero, but only marginally, which is analogous to borderline statistical significance at the 5% level. The frequentist check of BMA reported in Figure 6.5 shows statistical significance at the 10% level (and p-values larger than 0.5 for the other three country-level variables). We conclude that there seems to be a positive association between income and the elasticity of intertemporal substitution; the economic significance of this association is examined at the end of this section.

Figure 6.5: Posterior coefficient distributions for country characteristics



Notes: The figure depicts the densities of the regression parameters encountered in different regressions (with different subsets of control variables on the right-hand side). For example, the regression coefficient for *GDP per capita* is positive in almost all models, irrespective of the control variables included. The most common value of the coefficient is approximately 0.13. On the other hand, the coefficient for *Rule of law* is negative in one half of the models and positive in the other half, depending on which control variables are included. The most common value is 0.

As a next step we add the variable *stock market participation* to the model, which reduces the number of countries to 28—the ones for which information on stock market participation is available—and we label them "core countries." We are especially interested in the effect the new variable has on the estimated EIS, but we also examine the robustness of our results compared with the case where data for all countries were included. Even though this new BMA estimation includes far fewer countries, it only loses about 270 observations, because most studies estimate the EIS using data from the core countries.

The results of the BMA estimation with *stock market participation* are reported in Table 6.2; more details and diagnostics are available in Table 6.10 and Figure 6.8 in the Appendix. Concerning method characteristics, there are several changes compared with the estimation using all countries. First, it matters for the reported EIS whether the assumed utility function allows for non-separabilities between durable and non-durable consumption goods: allowing for non-separabilities is associated with larger estimated elasticities. Nevertheless, the variable has a posterior inclusion probability of only 0.54 and is not statistically significant in the frequentist check. Second, the posterior inclusion probability of the variable *exact Euler* drops to 0.29, so it seems to be less important when only the core countries are considered. Third, our results for the core countries suggest that highly cited studies report smaller estimates of the elasticity. But again, the corresponding variable has a posterior inclusion probability of only 0.6, and it is not significant in the frequentist check. Moreover, the posterior inclusion probability for this variable decreases sharply below 0.5 when we exclude the most cited study, Hall (1988), who reports small estimates.

Concerning the country-level variables, in the new BMA estimation we find a smaller posterior mean for the coefficient corresponding to *GDP per capita*; the variable also loses statistical significance in the frequentist check (nevertheless, the decrease in the posterior mean may reflect the positive correlation between *GDP per capita* and *stock market participation* of 0.54). The results concerning the remaining three variables do not change much, and the variables still appear to be quite unimportant. In contrast, the newly included *stock market participation* is positively

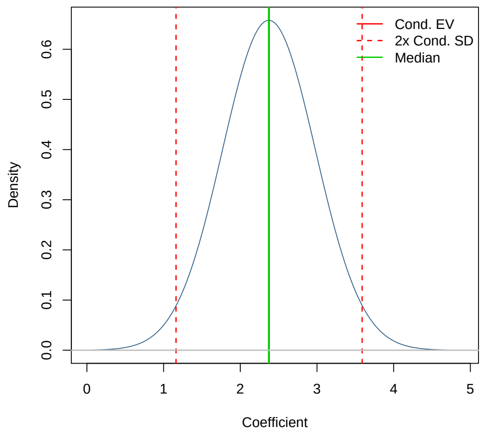
Table 6.2: Explaining the differences in the estimates of the EIS, core countries

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of the EIS	Post. mean	Post. std. dev.	PIP	Coef.	Std. er.	p-value
Country characteristics						
Stock market partic.	2.376	0.607	1.000	2.221	0.542	0.000
GDP per capita	0.080	0.137	1.000	0.116	0.138	0.405
Credit availability	-0.008	0.094	1.000	-0.003	0.122	0.982
Real interest	0.005	0.022	1.000	0.010	0.024	0.680
Rule of law	-0.283	0.193	1.000	-0.296	0.206	0.163
Utility						
Epstein-Zin	0.036	0.110	0.115			
Habits	-0.004	0.034	0.019			
Nonsep. durables	0.240	0.244	0.540	0.471	0.276	0.100
Nonsep. public	0.000	0.015	0.009			
Nonsep. tradables	0.004	0.042	0.016			
Data						
No. of households	-0.001	0.005	0.022			
No. of years	-0.248	0.059	0.996	-0.226	0.059	0.001
Average year	-0.025	0.860	0.010			
Micro data	-0.001	0.022	0.015			
Annual data	0.001	0.012	0.012			
Monthly data	0.141	0.166	0.506	0.326	0.054	0.000
Design						
Quasipanel	-0.107	0.191	0.273			
Inverse estimation	0.575	0.073	1.000	0.598	0.097	0.000
Asset holders	0.210	0.208	0.558	0.372	0.143	0.015
First lag instrument	0.002	0.019	0.022			
No year dummies	-0.007	0.066	0.021			
Income	-0.001	0.012	0.012			
Taste shifters	0.000	0.008	0.010			
Variable definition						
Total consumption	0.416	0.103	0.993	0.409	0.142	0.008
Food	0.016	0.080	0.057			
Stock return	-0.322	0.097	0.974	-0.358	0.158	0.032
Capital return	-0.224	0.164	0.714	-0.331	0.051	0.000
Estimation						
Exact Euler	0.067	0.114	0.287			
ML	-0.022	0.082	0.086			
TSLS	-0.002	0.021	0.022			
OLS	0.394	0.136	0.957	0.385	0.181	0.044
Publication						
Publication year	-0.074	1.288	0.012			
Citations	-0.052	0.048	0.595	-0.089	0.055	0.117
Top journal	0.529	0.104	1.000	0.567	0.103	0.000
Impact	0.000	0.004	0.016			
Constant	0.892	NA	1.000	-0.220	1.427	0.878
Observations	2,254			2,254		

Notes: EIS = elasticity of intertemporal substitution. PIP = posterior inclusion probability. Country characteristics are always included in all models of the BMA. In the frequentist check we only include method characteristics with PIP > 0.5. Standard errors in the frequentist check are clustered at the country level. More details on the BMA estimation are available in Table 6.10 and Figure 6.8.

associated with the estimated elasticities, as we can see from Figure 6.6. The regression parameter for this variable is positive in virtually all regressions in which the variable is included. Also, in the frequentist check the variable is highly statistically significant, with a p-value below 0.001. Our results thus suggest that households in countries with high stock market participation tend to be more willing or able to substitute consumption intertemporally.

Figure 6.6: Posterior coefficient distribution for *stock market participation*



Notes: The figure depicts the densities of the regression parameters encountered in different regressions (with different subsets of control variables on the right-hand side).

But is the effect of *stock market participation* economically important? The estimated posterior mean for the regression coefficient corresponding to the variable is 2.4, so that an increase in stock market participation of 10 percentage points is associated with an increase in the EIS of 0.24; an important difference according to the simulation shown in Figure 6.1. In Table 6.3 we compute what happens to the estimated elasticity if the value of a country-level characteristic changes from its sample minimum to its sample maximum ("maximum effect") and if the value increases by one standard deviation ("standard-deviation effect"). For variables *GDP per capita*, *credit availability*, *real interest*, and *rule of law*, we prefer to use the coefficients from the BMA estimation with all countries; for the variable *stock market participation* we have to use the value from the estimation with the core countries only. Out of the five country-level variables, *stock market participation* has the largest effect, followed by *GDP per capita*. The other variables do not seem to matter much. The maximum effect of changes in *stock market participation* is a whopping 0.93; the standard-deviation effect is 0.14, which can also make a difference to the results of structural models, as shown in Figure 6.1.

Table 6.3: The economic significance of differences in country characteristics

Variable	Maximum effect	Std. dev. effect
Stock market partic.	0.931	0.141
GDP per capita	0.683	0.088
Credit availability	-0.119	-0.020
Real interest	-0.265	-0.019
Rule of law	-0.087	-0.012

Notes: The table depicts the predicted effects of increases in the variables on the EIS estimates based on the BMA results (the specification with core countries for *stock market participation*; the specification with all countries for the other variables). Maximum effect = an increase from sample minimum to sample maximum. Std. dev. effect = a one-standard-deviation increase.

6.5 Robustness Checks

In this section we evaluate the robustness of our findings by employing different variants of the BMA specification with the core countries—that is, including the variable *Stock market participation*. First, we run a BMA estimation in which country-level variables are treated in the same way as method variables; in other words, different models may or may not include country-level variables, in contrast to the previous analysis, in which country-level variables were included in all models. Table 6.4 provides the results (here we do not report results for variables with posterior inclusion probability below 0.5), and more details and diagnostics are available in Table 6.11 and Figure 6.9 in the Appendix.

In this estimation the posterior inclusion probabilities for country-level variables are not necessarily 1, and indeed the probabilities for all variables except *stock market participation* are lower than 0.5, which means that these variables do not help us explain the variation in the reported elasticities once the characteristics of methodology are taken into account. In contrast, the posterior inclusion probability of *Stock market participation* is 0.92, which would be characterized as "substantial" in the guidelines for the interpretation of the posterior inclusion probability by Eicher *et al.* (2011). Moreover, in the frequentist check the variable is statistically significant at the 1% level.

The regression parameter for *stock market participation* estimated by BMA is now lower than in the previous case, but still implies an important effect on the estimated EIS: an increase in stock market participation of 10 percentage points is associated with an increase in the estimated elasticity of 0.18. Concerning the method variables, the results of the robustness check are similar to the baseline case, where the country-level variables are included in all models, but a few differences emerge. First, the data frequency does not seem to be important for the estimated EIS when country and method variables are treated in the same way. Second, the results suggest that estimating the exact Euler equation, instead of the log-linearized version, tends to deliver larger elasticities—we reported the same finding for the BMA estimation with all countries (that is, excluding *stock market participation*). Third, according to this robustness check the number of study citations is not associated with the magnitude of the reported elasticity.

The second robustness check involves different priors for the BMA estimation. Now we use the priors that are advocated by Eicher *et al.* (2011) because they typically perform well in forecasting exercises: the unit information g-prior (the prior provides the same amount of information as one observation) and the uniform model prior (each model has the same probability). As we have noted, BMA runs many regressions with different combinations of the explanatory variables on the right-hand side and not all of the variables have to be included. It follows that models of size 15—the number of explanatory variables divided by two—are most common. If each model has the same probability, with the uniform model prior we implicitly impose the prior that the "true" model explaining the differences in the reported elasticities has 15 explanatory variables, which is apparent from Figure 6.10 in the Appendix.

Table 6.4: Robustness check: no fixed variables

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of the EIS	Post. mean	Post. std. dev.	PIP	Coef.	Std. er.	p-value
Stock market partic.	1.775	0.736	0.917	2.128	0.613	0.002
GDP per capita	0.000	0.010	0.008	0.060	0.166	0.721
Credit availability	-0.002	0.016	0.021	0.040	0.129	0.760
Real interest	0.000	0.002	0.008	-0.004	0.026	0.879
Rule of law	-0.013	0.062	0.053	-0.290	0.238	0.234
Inverse estimation	0.563	0.078	1.000	0.535	0.146	0.001
Top journal	0.502	0.103	1.000	0.418	0.074	0.000
Total consumption	0.449	0.095	0.999	0.439	0.101	0.000
No. of years	-0.255	0.056	0.999	-0.232	0.050	0.000
Stock return	-0.340	0.088	0.990	-0.341	0.139	0.022
OLS	0.438	0.120	0.986	0.521	0.148	0.002
Capital return	-0.231	0.160	0.735	-0.282	0.054	0.000
Asset holders	0.277	0.210	0.694	0.404	0.115	0.002
Exact Euler	0.138	0.144	0.522	0.283	0.226	0.221
Constant	0.746	NA	1.000	0.105	1.634	0.950
Observations	2,254			2,254		

Notes: PIP = posterior inclusion probability. Country characteristics and method variables are treated in the same way in the BMA estimation. Results for method characteristics with PIP $<$ 0.5 are not reported. Standard errors in the frequentist check are clustered at the country level. More details on the BMA estimation are available in Table 6.11 and Figure 6.9.

That is why for the baseline estimation we prefer the random model prior, which gives each model *size* the same prior probability and reflects the fact that we know little ex ante about how many variables should be included in the model. The results of the robustness check are reported in Table 6.5 and for both country-level and method variables they are virtually identical to the baseline case.

Finally, in the third robustness check we use different proxies for liquidity constraints and institutions. Instead of the measure of credit availability reported in the Global Competitiveness Report we now employ the measure of financial reform published by the IMF; instead of perceptions of the rule of law in society we employ the measure of generalized trust developed by Bjoernskov & Meon (2013). The result concerning *stock market participation* holds: the variable is positively and strongly associated with the elasticity of intertemporal substitution. The other variables are less important, even though *GDP per capita* and *Financial reform* yield statistical significance at the 10% level in the frequentist check of the BMA estimation. Concerning the method variables, the results are close to the baseline case, with the exception of data frequency, which seems to be unimportant here, similarly to the

Table 6.5: Robustness check: priors according to Eicher *et al.* (2011)

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of the EIS	Post. mean	Post. std. dev.	PIP	Coef.	Std. er.	p-value
Stock market partic.	2.328	0.598	1.000	2.221	0.542	0.000
GDP per capita	0.082	0.137	1.000	0.116	0.138	0.405
Credit availability	-0.018	0.095	1.000	-0.003	0.122	0.982
Real interest	0.007	0.022	1.000	0.010	0.024	0.680
Rule of law	-0.258	0.192	1.000	-0.296	0.206	0.163
Inverse estimation	0.594	0.070	1.000	0.598	0.097	0.000
Top journal	0.554	0.101	1.000	0.567	0.103	0.000
Stock return	-0.345	0.081	0.998	-0.358	0.158	0.032
Total consumption	0.416	0.098	0.998	0.409	0.142	0.008
No. of years	-0.247	0.059	0.998	-0.226	0.059	0.001
OLS	0.383	0.127	0.969	0.385	0.181	0.044
Capital return	-0.305	0.128	0.921	-0.331	0.051	0.000
Asset holders	0.294	0.192	0.771	0.372	0.143	0.015
Citations	-0.067	0.045	0.762	-0.089	0.055	0.117
Nonsep. durables	0.331	0.231	0.738	0.471	0.276	0.100
Monthly data	0.193	0.165	0.641	0.326	0.054	0.000
Constant	1.199	NA	1.000	-0.220	1.427	0.878
Observations	2,254			2,254		

Notes: PIP = posterior inclusion probability. In this specification we employ the priors suggested by Eicher *et al.* (2011), who recommend using the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation). Results for method characteristics with PIP < 0.5 are not reported. Standard errors in the frequentist check are clustered at the country level. More details on the BMA estimation are available in Table 6.12 and Figure 6.10.

first robustness check and the BMA estimation with all countries.

As we have noted, for all analyses in the paper we exclude estimates of the EIS larger than 10 in absolute value. It is necessary to exclude outliers because the inverse method of estimation used by some researchers can yield implausible estimates of the elasticity—even larger than 100 in absolute value. Because with the asset return on the left-hand side the researcher estimates the inverse of the EIS (the coefficient of relative risk aversion under the typical power utility), imprecise estimation may yield a coefficient close to zero and imply that the EIS is close to infinity. The threshold of 10 is arbitrary, but we get very similar results with the threshold set to 1, 5, 20, and 100. Moreover, the results are also similar when we include all estimates of the EIS and employ the robust estimator developed by Verardi & Croux (2009) for the frequentist check. As far as we know, a variant of robust estimation is not yet available for the BMA framework.

Table 6.6: Robustness check: alternative proxies for liquidity constraints and institutions

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of the EIS	Post. mean	Post. std. dev.	PIP	Coef.	Std. er.	p-value
Stock market partic.	2.399	0.609	1.000	2.342	0.848	0.011
GDP per capita	0.137	0.142	1.000	0.198	0.114	0.095
Financial reform	-0.692	0.307	1.000	-0.777	0.394	0.060
Real interest	0.025	0.023	1.000	0.023	0.032	0.493
Trust	-0.006	0.005	1.000	-0.005	0.004	0.257
Inverse estimation	0.577	0.075	1.000	0.627	0.103	0.000
Top journal	0.543	0.104	1.000	0.602	0.114	0.000
Total consumption	0.423	0.100	0.996	0.416	0.147	0.009
No. of years	-0.236	0.061	0.991	-0.228	0.058	0.001
OLS	0.412	0.126	0.976	0.443	0.189	0.028
Stock return	-0.303	0.101	0.961	-0.299	0.136	0.037
Asset holders	0.299	0.211	0.728	0.406	0.130	0.005
Citations	-0.063	0.049	0.682	-0.093	0.057	0.119
Capital return	-0.182	0.168	0.596	-0.265	0.061	0.000
Nonsep. durables	0.257	0.247	0.570	0.465	0.273	0.101
Constant	-0.440	NA	1.000	-0.797	1.093	0.473
Observations	2,254			2,254		

Notes: PIP = posterior inclusion probability. In this specification we replace *Credit availability* with *Financial reform* and *Rule of law* with *Trust*. Results for method characteristics with PIP < 0.5 are not reported. Standard errors in the frequentist check are clustered at the country level. More details on the BMA estimation are available in Table 6.13 and Figure 6.11.

6.6 Concluding Remarks

We present a quantitative survey of estimates of the elasticity of intertemporal substitution in what we believe is the largest meta-analysis conducted in economics. We collect 2,735 estimates from 169 published studies and find that the mean elasticity is 0.5, but that the estimates vary greatly across countries and methods. We use Bayesian model averaging to explore country-level heterogeneity while controlling for 30 variables that reflect different techniques used in the estimation of the elasticity. We find that households in countries with higher income per capita and higher stock market participation show larger values of the EIS. Thus, using a unique cross-country data set we corroborate the micro-level findings of Blundell *et al.* (1994) and Attanasio & Browning (1995), who report a larger elasticity for richer households, and Mankiw & Zeldes (1991) and Vissing-Jorgensen (2002), who find a larger EIS for asset holders than for other households. Our results also suggest that researchers obtain systematically larger estimates of the EIS when they estimate the parameter using a sub-sample of rich households or asset holders.

Rich households substitute consumption across time periods more easily because necessities, which are difficult to substitute intertemporally, constitute a smaller fraction of their consumption bundle in comparison with poor households. Moreover, the opportunities for intertemporal substitution for households in developing countries may be restricted by subsistence requirements (Ogaki *et al.* 1996). Concerning asset holders, Vissing-Jorgensen (2002) points out that the consumption Euler equation need not be valid for households that do not participate in asset markets, leading to estimates of the EIS close to zero. Another possible explanation is that exposure to financial markets, especially the stock market, may make households more forward-looking and willing to substitute consumption in response to changes in expected asset returns.

Several aspects of methodology affect the reported elasticities in a systematic way. For example, the definition of the utility function is important, especially whether researchers allow for non-separabilities between durable and non-durable consumption goods. The size of the data set matters for the estimated elasticities as well. Further, when researchers use asset returns as the response variable and estimate the inverse of the EIS, the implied elasticity tends to be substantially larger—on average by about 0.5 compared to the case where consumption growth is used as the response variable. The definition of consumption growth (total consumption, non-durables, or food expenditure) and asset return (bond, stock, or capital return) is also important. Ignoring the presence of endogeneity typically leads to overestimation of the elasticity. Finally, the top five general interest journals in economics tend to publish substantially larger estimates than other journals, which may reflect unobserved aspects of study quality.

An important issue that we do not discuss in this paper is publication selection bias. Several commentators have suggested that in empirical economics statistically insignificant results tend to be underreported and that the resulting mean estimate observed in the literature may be biased (DeLong & Lang 1992; Card & Krueger 1995; Ashenfelter & Greenstone 2004; Stanley 2005). We analyze publication selection bias in the EIS literature in a companion paper, Havranek (2013), and believe that while such bias can affect the mean reported elasticity, it is not related to country-level heterogeneity in the EIS.

References

Abiad, A., E. Detragiache, & T. Tressel (2010): "A New Database of Financial Reforms." *IMF Staff Papers* **57(2)**: pp. 281–302.

Ai, H. (2010): "Information Quality and Long-Run Risk: Asset Pricing Implications." *Journal of Finance* **65(4)**: pp. 1333–1367.

Ashenfelter, O. & M. Greenstone (2004): "Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias." *American Economic Review* **94(2)**: pp. 454–460.

Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* **6(4)**: pp. 453–470.

Atkeson, A. & M. Ogaki (1996): "Wealth-varying intertemporal elasticities of substitution: Evidence from panel and aggregate data." *Journal of Monetary Economics* **38(3)**: pp. 507–534.

Attanasio, O. P. (1995): "The intertemporal allocation of consumption: theory and evidence." *Carnegie-Rochester Conference Series on Public Policy* **42(1)**: pp. 39–56.

Attanasio, O. P., J. Banks, & S. Tanner (2002): "Asset Holding and Consumption Volatility." *Journal of Political Economy* **110(4)**: pp. 771–792.

Attanasio, O. P. & M. Browning (1995): "Consumption over the Life Cycle and over the Business Cycle." *American Economic Review* **85(5)**: pp. 1118–37.

Attanasio, O. P. & G. Weber (1993): "Consumption Growth, the Interest Rate and Aggregation." *Review of Economic Studies* **60(3)**: pp. 631–49.

Attanasio, O. P. & G. Weber (1995): "Is Consumption Growth Consistent with

Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* **103(6)**: pp. 1121–57.

Barro, R. J. (2009): "Rare Disasters, Asset Prices, and Welfare Costs." *American Economic Review* **99(1)**: pp. 243–64.

Bayoumi, T. A. (1993): "Financial Deregulation and Consumption in the United Kingdom." *The Review of Economics and Statistics* **75(3)**: pp. 536–39.

Bjoernskov, C. & P.-G. Meon (2013): "Is trust the missing root of institutions, education, and development?" *Public Choice* **157**: pp. 641–669.

Blundell, R., M. Browning, & C. Meghir (1994): "Consumer Demand and the Life-Cycle Allocation of Household Expenditures." *Review of Economic Studies* **61(1)**: pp. 57–80.

Browning, M. & A. Lusardi (1996): "Household Saving: Micro Theories and Micro Facts." *Journal of Economic Literature* **34(4)**: pp. 1797–1855.

Campbell, J. Y. (1999): "Asset Prices, Consumption, and the Business Cycle." In J. B. Taylor & M. Woodford (editors), "Handbook of Macroeconomics," volume 1 of *Handbook of Macroeconomics*, chapter 19, pp. 1231–1303. Elsevier.

Campbell, J. Y. & N. G. Mankiw (1989): "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence." In "NBER Macroeconomics Annual 1989, Volume 4," NBER Chapters, pp. 185–246. National Bureau of Economic Research, Inc.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-Analysis." *American Economic Review* **85(2)**: pp. 238–43.

Carroll, C. D. (2001): "Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation)." *Advances in Macroeconomics* **1**: p. 6.

Chari, V. V., P. J. Kehoe, & E. R. McGrattan (2002): "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?" *Review of Economic Studies* **69(3)**: pp. 533–63.

Chetty, R., A. Guren, D. Manoli, & A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* **101(3)**: pp. 471–75.

Ciccone, A. & M. Jarocinski (2010): "Determinants of Economic Growth: Will Data Tell?" *American Economic Journal: Macroeconomics* **2(4)**: pp. 222–46.

Colacito, R. & M. M. Croce (2011): "Risks for the Long Run and the Real Exchange Rate." *Journal of Political Economy* **119(1)**: pp. 153–181.

Crossley, T. F. & H. W. Low (2011): "Is the elasticity of intertemporal substitution constant?" *Journal of the European Economic Association* **9(1)**: pp. 87–105.

DeLong, J. B. & K. Lang (1992): "Are All Economic Hypotheses False?" *Journal of Political Economy* **100(6)**: pp. 1257–72.

Disdier, A.-C. & K. Head (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics* **90(1)**: pp. 37–48.

Doucouliagos, H. & T. D. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Epstein, L. G. & S. E. Zin (1989): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework." *Econometrica* **57(4)**: pp. 937–69.

Feldkircher, M. & S. Zeugner (2009): "Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging." *IMF Working Papers 09/202*, International Monetary Fund.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

Giannetti, M. & Y. Koskinen (2010): "Investor Protection, Equity Returns, and

Financial Globalization." *Journal of Financial and Quantitative Analysis* **45(01)**: pp. 135–168.

Gruber, J. (2006): "A Tax-Based Estimate of the Elasticity of Intertemporal Substitution." *NBER Working Papers 11945*, National Bureau of Economic Research, Inc.

Guiso, L., P. Sapienza, & L. Zingales (2004): "The Role of Social Capital in Financial Development." *American Economic Review* **94(3)**: pp. 526–556.

Guiso, L., P. Sapienza, & L. Zingales (2008): "Trusting the Stock Market." *Journal of Finance* **63(6)**: pp. 2557–2600.

Hall, R. E. (1988): "Intertemporal Substitution in Consumption." *Journal of Political Economy* **96(2)**: pp. 339–57.

Havranek, T. (2013): "Publication Bias in Measuring Intertemporal Substitution." *Working paper*, Czech National Bank and Charles University in Prague. Available at `meta-analysis.cz/eis`.

Havranek, T. & M. Rusnak (2013): "Transmission Lags of Monetary Policy: A Meta-Analysis." *International Journal of Central Banking* **9(4)**: pp. 39–76.

House, C. L. & M. D. Shapiro (2006): "Phased-In Tax Cuts and Economic Activity." *American Economic Review* **96(5)**: pp. 1835–1849.

Irsova, Z. & T. Havranek (2013): "Determinants of Horizontal Spillovers from FDI: Evidence from a Large Meta-Analysis." *World Development* **42(1)**: pp. 1–15.

Jin, K. (2012): "Industrial Structure and Capital Flows." *American Economic Review* **102(5)**: pp. 2111–46.

Ley, E. & M. F. Steel (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions." *Journal of Applied Econometrics* **24(4)**: pp. 651–674.

Mankiw, N. G. & S. P. Zeldes (1991): "The Consumption of Stockholders and Non-Stockholders." *NBER Working Papers 3402*, National Bureau of Economic

Research, Inc.

Moeltner, K. & R. Woodward (2009): "Meta-Functional Benefit Transfer for Wetland Valuation: Making the Most of Small Samples." *Environmental & Resource Economics* **42(1)**: pp. 89–108.

Moral-Benito, E. (2012): "Determinants of Economic Growth: A Bayesian Panel Data Approach." *The Review of Economics and Statistics* **94(2)**: pp. 566–579.

Mulligan, C. B. (2002): "Capital, Interest, and Aggregate Intertemporal Substitution." *NBER Working Papers 9373*, National Bureau of Economic Research, Inc.

Nelson, J. & P. Kennedy (2009): "The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment." *Environmental and Resource Economics* **42**: pp. 345–377.

Ogaki, M., J. D. Ostry, & C. M. Reinhart (1996): "Saving Behavior in Low- and Middle-Income Developing Countries: A Comparison." *IMF Staff Papers* **43(1)**: pp. 38–71.

Ogaki, M. & C. M. Reinhart (1998): "Measuring Intertemporal Substitution: The Role of Durable Goods." *Journal of Political Economy* **106(5)**: pp. 1078–1098.

Piazzesi, M., M. Schneider, & S. Tuzel (2007): "Housing, Consumption and Asset Pricing." *Journal of Financial Economics* **83(3)**: pp. 531–569.

Porta, R. L., F. L. de Silanes, A. Shleifer, & R. W. Vishny (1998): "Law and Finance." *Journal of Political Economy* **106(6)**: pp. 1113–1155.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* **92**: pp. 179–191.

Rieger, M. O., M. Wang, & T. Hens (2011): "Prospect Theory around the World." *Discussion Papers 2011/19*, Department of Finance and Management Science, Norwegian School of Economics.

Rudebusch, G. D. & E. T. Swanson (2012): "The Bond Premium in a DSGE

Model with Long-Run Real and Nominal Risks." *American Economic Journal: Macroeconomics* **4(1)**: pp. 105–43.

Smets, F. & R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* **97(3)**: pp. 586–606.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-Analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Trabandt, M. & H. Uhlig (2011): "The Laffer Curve Revisited." *Journal of Monetary Economics* **58(4)**: pp. 305–327.

Verardi, V. & C. Croux (2009): "Robust regression in Stata." *Stata Journal* **9(3)**: pp. 439–453.

Vissing-Jorgensen, A. (2002): "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *Journal of Political Economy* **110(4)**: pp. 825–853.

Wang, M., M. O. Rieger, & T. Hens (2011): "How Time Preferences Differ: Evidence from 45 Countries." *Discussion Papers 2011/18*, Department of Finance and Management Science, Norwegian School of Economics.

Wieland, V., T. Cwik, G. J. Müller, S. Schmidt, & M. Wolters (2012): "A new comparative approach to macroeconomic modeling and policy analysis." *Journal of Economic Behavior & Organization* **83(3)**: pp. 523–541.

Wirjanto, T. S. (1995): "Aggregate Consumption Behaviour and Liquidity Constraints: The Canadian Evidence." *Canadian Journal of Economics* **28(4b)**: pp. 1135–52.

6.A Summary Statistics

Table 6.8: Description and summary statistics of regression variables

Variable	Description	Mean	Std. dev.
EIS	Estimate of the elasticity of intertemporal substitution (response variable).	0.492	1.298
Country characteristics			
Stock market partic.	The fraction of households participating in the domestic stock market (source: Giannetti & Koskinen 2010).	0.246	0.059
GDP per capita	Gross domestic product per capita at purchasing-power-adjusted 2005 dollars (source: Penn World Tables).	9.804	0.658
Credit availability	The ease of access to loans (source: The Global Competitiveness Report, `www.weforum.org`).	3.523	0.547
Financial reform	The IMF's financial reform index (source: Abiad *et al.* 2010).	0.691	0.197
Real interest	The lending interest rate adjusted for inflation as measured by the GDP deflator (source: World Development Indicators).	4.448	3.954
Rule of law	The extent to which agents have confidence in the rules of society, and in particular the quality of contract enforcement (source: World Bank Global Governance Indicators).	1.404	0.611
Trust	Perceptions of general trust in society (source: Bjoernskov & Meon 2013).	39.09	9.543
Method characteristics			
Utility			
Epstein-Zin	=1 if the estimation differentiates between the EIS and the coefficient of relative risk aversion.	0.053	0.224
Habits	=1 if habits in consumption are assumed.	0.040	0.196
Nonsep. durables	=1 if the model allows for nonseparability between durables and nondurables.	0.041	0.199
Nonsep. public	=1 if the model allows for nonseparability between private and public consumption.	0.044	0.206
Nonsep. tradables	=1 if the model allows for nonseparability between tradables and nontradables.	0.046	0.210
Data			
No. of households	The logarithm of the number of cross-sectional units used in the estimation (households, cohorts, countries).	1.103	2.384
No. of years	The logarithm of the number of years of the data period used in the estimation.	3.184	0.570
Average year	The logarithm of the average year of the data period.	7.590	0.006
Micro data	=1 if the coefficient comes from a micro-level estimation.	0.187	0.390
Annual data	=1 if the data frequency is annual.	0.328	0.469
Monthly data	=1 if the data frequency is monthly.	0.097	0.296
Design			
Quasipanel	=1 if quasipanel (synthetic cohort) data are used.	0.053	0.224
Inverse estimation	=1 if the rate of return is the response variable in the estimation.	0.317	0.465
Asset holders	=1 if the estimate is related to the rich or asset holders.	0.054	0.226
First lag instrument	=1 if the first lags of variables are included among the instruments.	0.305	0.460

Continued on next page

Description and summary statistics of regression variables (continued)

Variable	Description	Mean	Std. dev.
No year dummies	=1 if year dummies are omitted in micro studies using the Panel Study of Income Dynamics.	0.030	0.171
Income	=1 if income is included in the specification.	0.241	0.428
Taste shifters	The logarithm of the number of controls for taste shifters.	0.117	0.452
Variable definition			
Total consumption	=1 if total consumption is used in the estimation.	0.203	0.402
Food	=1 if food is used as a proxy for nondurables.	0.059	0.235
Stock return	=1 if the rate of return is measured as the stock return.	0.189	0.392
Capital return	=1 if the rate of return is measured as the return on capital.	0.113	0.317
Estimation			
Exact Euler	=1 if the exact Euler equation is estimated.	0.238	0.426
ML	=1 if maximum likelihood methods are used for the estimation.	0.049	0.216
TSLS	=1 if two-stage least squares are used for the estimation.	0.338	0.473
OLS	=1 if ordinary least squares are used for the estimation.	0.104	0.306
Publication			
Publication year	The logarithm of the year of publication of the study.	7.601	0.004
Citations	The logarithm of the number of per-year citations of the study in Google Scholar.	2.024	1.256
Top journal	=1 if the study was published in one of the top five journals in economics.	0.207	0.405
Impact	The recursive RePEc impact factor of the outlet.	1.089	1.535

Notes: Method characteristics are collected from published studies estimating the elasticity of intertemporal substitution. The list of studies is available in the online appendix at `meta-analysis.cz/substitution`.

Table 6.7: Meta-analyses of the EIS for individual countries

Country	Mean EIS	Std. err. of the mean	Estimates
Argentina	-0.171	0.221	12
Australia	0.362	0.160	32
Austria	3.149	1.876	6
Belgium	0.677	0.390	10
Brazil	0.107	0.093	19
Burma	0.439	0.042	4
Canada	0.389	0.110	91
Chile	0.137	0.077	7
China	0.530	0.234	5
Colombia	0.158	0.078	8
Denmark	0.488	0.588	7
Finland	0.185	0.320	46
France	-0.034	0.153	44
Germany	0.080	0.163	39
Greece	0.561	0.291	18
Hong Kong	0.099	0.017	33
Iceland	0.352	0.367	4
India	0.515	0.090	5
Indonesia	0.102	0.160	8
Ireland	1.739	0.778	7
Israel	0.235	0.033	65
Italy	0.290	0.162	33
Japan	0.893	0.243	109
Kenya	1.228	0.481	7
Korea	0.423	0.219	32
Malaysia	0.173	0.161	11
Mexico	0.158	0.053	12
Netherlands	0.027	0.221	31
New Zealand	2.206	0.269	4
Norway	-0.386	0.583	4
Pakistan	0.100	0.203	6
Philippines	-0.026	0.111	9
Portugal	0.152	0.258	7
Singapore	0.120	0.131	7
Spain	0.504	0.107	44
Sri Lanka	0.033	0.159	8
Sweden	0.065	0.126	63
Switzerland	-0.434	0.201	31
Taiwan	1.549	1.421	7
Thailand	0.081	0.064	9
Turkey	0.314	0.133	12
UK	0.487	0.070	251
Uruguay	0.117	0.124	5
US	0.594	0.036	1429
Venezuela	0.157	0.093	6

Notes: The table shows mean estimates of the EIS in countries for which at least 4 estimates are reported in the literature. Estimates larger than 10 in absolute value are excluded.

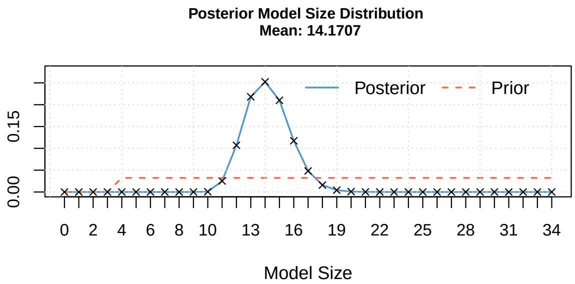
6.B Diagnostics of BMA

Table 6.9: Summary of BMA estimation, all countries

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
14.1707	$2 \cdot 10^6$	$1 \cdot 10^6$	8.14355 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
377,919	$1.7 \cdot 10^{10}$	0.0022%	96%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9999	2,526	random	BRIC
Shrinkage-Stats			
Av= 0.9996			

Notes: The "random" model prior refers to the beta-binomial prior advocated by Ley & Steel (2009): prior model probabilities are the same for all possible model sizes. We set the Zellner's g prior following Fernandez *et al.* (2001).

Figure 6.7: Model size and convergence, BMA with all countries



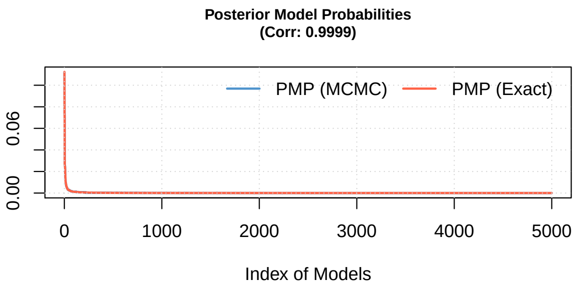


Table 6.10: Summary of BMA estimation, core countries

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
14.9218	$2 \cdot 10^6$	$1 \cdot 10^6$	8.464817 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
478, 214	$3.4 \cdot 10^{10}$	0.0014%	94%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9996	2, 254	random	BRIC
Shrinkage-Stats			
Av= 0.9996			

Notes: The "random" model prior refers to the beta-binomial prior advocated by Ley & Steel (2009): prior model probabilities are the same for all possible model sizes. We set the Zellner's g prior following Fernandez *et al.* (2001).

Figure 6.8: Model size and convergence, BMA with core countries

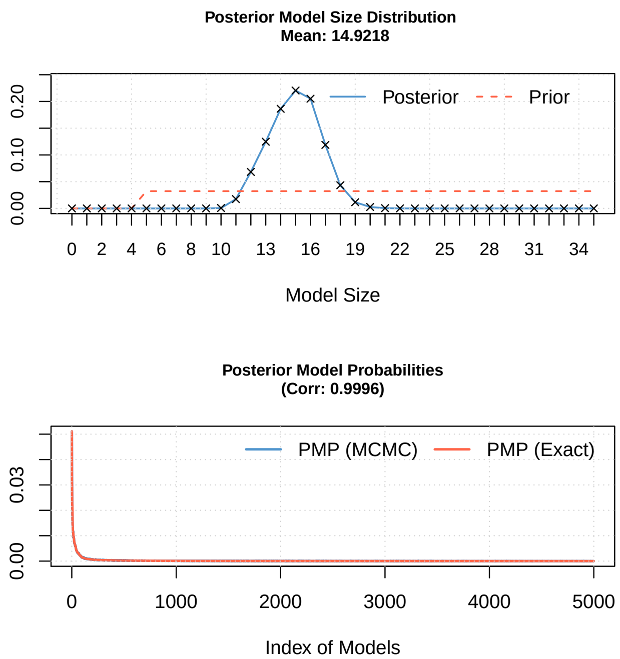


Table 6.11: Summary of BMA estimation, no fixed variables

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
10.9643	$2 \cdot 10^6$	$1 \cdot 10^6$	7.003633 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
387,615	$3.4 \cdot 10^{10}$	0.0011%	92%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9995	2,254	random	BRIC
Shrinkage-Stats			
Av= 0.9996			

Notes: The "random" model prior refers to the beta-binomial prior advocated by Ley & Steel (2009): prior model probabilities are the same for all possible model sizes. We set the Zellner's g prior following Fernandez *et al.* (2001).

Figure 6.9: Model size and convergence, BMA with no fixed variables

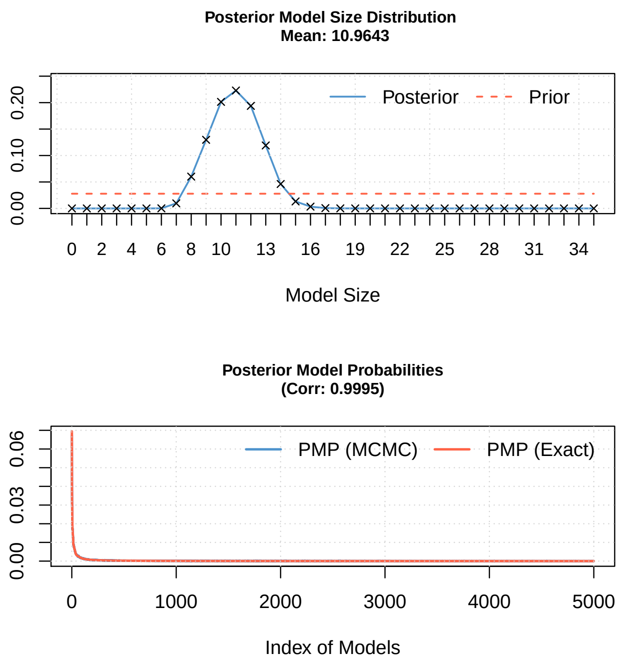


Table 6.12: Summary of BMA estimation, priors according to Eicher *et al.* (2011)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
16.3370	$2 \cdot 10^6$	$1 \cdot 10^6$	8.44965 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
497,193	$3.4 \cdot 10^{10}$	0.0014%	90%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9994	2,254	uniform	UIP
Shrinkage-Stats			
Av= 0.9996			

Notes: In this specification we employ the priors suggested by Eicher *et al.* (2011), who recommend using the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation).

Figure 6.10: Model size and convergence, BMA with priors according to Eicher *et al.* (2011)

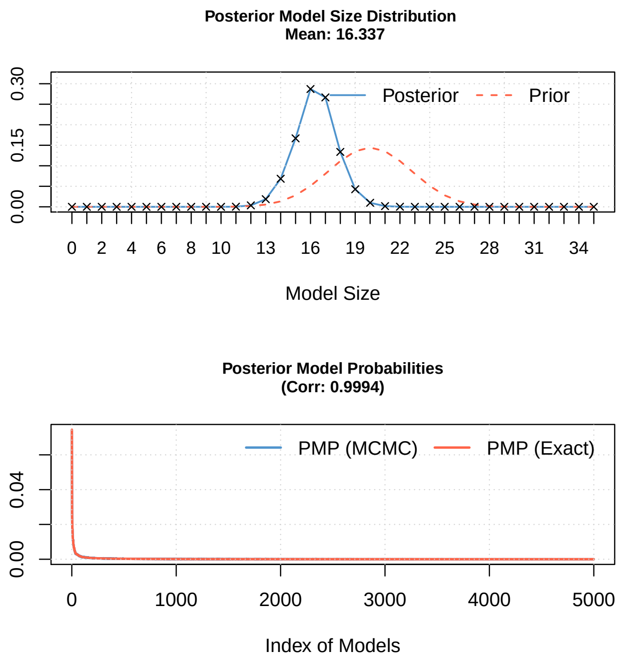
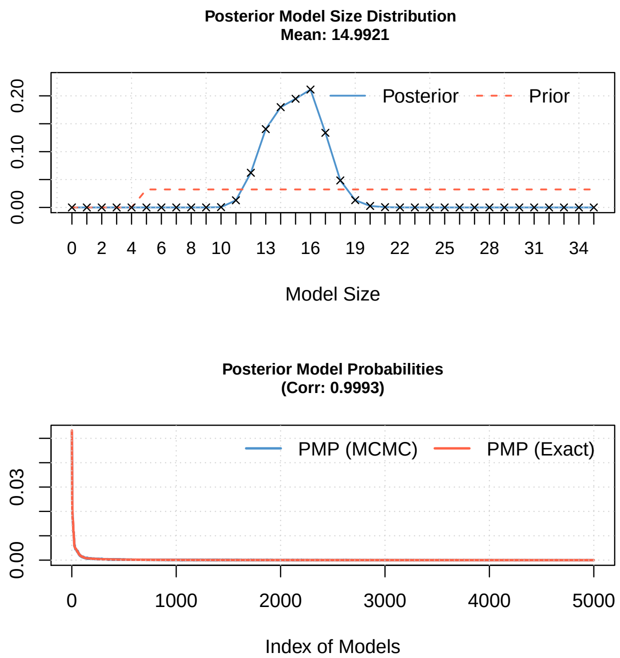


Table 6.13: Summary of BMA estimation, alternative proxies

Mean no. regressors 14.9921	*Draws* $2 \cdot 10^6$	*Burn-ins* $1 \cdot 10^6$	*Time* 8.557683 minutes
No. models visited 443,396	*Modelspace* $3.4 \cdot 10^{10}$	*Visited* 0.0013%	*Topmodels* 95%
Corr PMP 0.9993	*No. Obs.* 2,254	*Model Prior* random	*g-Prior* BRIC
Shrinkage-Stats Av= 0.9996			

Notes: The "random" model prior refers to the beta-binomial prior advocated by Ley & Steel (2009): prior model probabilities are the same for all possible model sizes. We set the Zellner's g prior following Fernandez *et al.* (2001).

Figure 6.11: Model size and convergence, BMA with alternative proxies



Chapter 7

Selective Reporting and the Social Cost of Carbon

Abstract

We examine potential selective reporting in the literature on the social cost of carbon (SCC) by conducting a meta-analysis of 809 estimates of the SCC reported in 101 studies. Our results indicate that estimates for which the 95% confidence interval includes zero are less likely to be reported than estimates excluding negative values of the SCC, which creates an upward bias in the literature. The evidence for selective reporting is stronger for studies published in peer-reviewed journals than for unpublished papers. We show that the findings are not driven by the asymmetry of confidence intervals surrounding the SCC and are robust to controlling for various characteristics of study design and to alternative definitions of confidence intervals. Our estimates of the mean reported SCC corrected for the selective reporting bias are imprecise and range between 0 and 130 USD per ton of carbon in 2010 prices for emission year 2015.

Keywords: Social cost of carbon, climate policy, integrated assessment models, meta-analysis, selective reporting, publication bias

JEL Codes: C83, Q54

This paper is a joint work with Tomas Havranek, Karel Janda, and David Zilberman. An online appendix with data, code, and a list of studies included in the meta-analysis is available at `meta-analysis.cz/scc`. Zuzana Irsova acknowledges support from the Grant Agency of Charles University in Prague (grant #558713); Tomas Havranek acknowledges support from the Czech Science Foundation (grant #P402/11/0948). The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013 under REA grant agreement number 609642. We thank Richard Tol for sending us his data set and are grateful to Jan Babecky, Jiri Schwarz, Diana Zigraiova, and seminar participants at Charles University for their helpful comments. The paper is currently under review in the Journal of Environmental Economics and Management.

7.1 Introduction

A key parameter for the formulation of climate policy is the social cost of carbon emissions. If the social cost of carbon was pinned down precisely, policy makers could use the parameter to set the optimal carbon tax. For this reason, dozens of researchers using different families of models have estimated the SCC—but their findings and the resulting policy implications vary greatly. Several previous studies have offered quantitative surveys of the literature (Tol 2005b; 2008; 2011; 2013b), focusing especially on the characteristics of study design that may influence the reported estimates, but no study has discussed nor tested for the potential selective reporting bias in the estimates of the social cost of carbon.

Selective reporting is the tendency of editors, referees, or authors themselves to prefer empirical estimates that are conclusive, have a particular sign supported by theory or intuition, or both. Also called the file-drawer problem or publication bias (we prefer the term selective reporting because the bias can be present in unpublished studies as well), it has been discussed in literature surveys since Rosenthal (1979). The problem of selective reporting is widely recognized in medical research, where many of the best journals now require prior registration of clinical trials as a necessary condition for any potential submission of results (Krakovsky 2004). In a similar vein, the American Economic Association has agreed to establish a registry of randomized controlled experiments to counter selective reporting (Siegfried 2012, p. 648).

Doucouliagos & Stanley (2013) conduct a large survey of meta-analyses (quantitative literature surveys) in economics and conclude that most fields suffer from selective reporting, which exaggerates the magnitude of the mean reported effect, and thus biases our inference from the literature. A recent survey among the members of the European Economic Association, Necker (2014), reveals that a third of economists in Europe admit that they have engaged in presenting empirical findings selectively so they confirm their arguments and in searching for control variables until they get a desired result. A meta-analysis by Havranek *et al.* (2012) indicates that 40% of the estimates of the price elasticity of gasoline demand end up hidden in researchers' file drawers because of an unintuitive sign or statistical insignificance; this selective reporting exaggerates the mean reported price elasticity twofold.

Several studies examine selective reporting in the context of climate change research. The problem is widely discussed in phenology (Both *et al.* 2004; Gienapp *et al.* 2007; Menzel *et al.* 2006), and the evidence suggests that while selective reporting is a minor issue in multi-species studies, positive results from single-species studies are reported more often than neutral results (Parmesan 2007). Maclean & Wilson (2011) conduct a meta-analysis of the relation between climate change and extinction risk and find mixed results concerning selective reporting, with evidence for the bias among estimates of extinction risk, but no bias among estimates of high extinction risk. Michaels (2008) examines 166 papers on climate change published in Science and Nature and argues that there is substantial evidence for selective reporting. Swanson (2013) indicates that many of the current model simulations of climate change are inconsistent with the observed changes in air temperature and the frequency of monthly temperature extremes, which might be due to selective reporting. In contrast, Darling & Côté (2008) investigate the relationship between climate change and biodiversity loss and find no evidence of selective reporting, and Massad & Dyer (2010) find no signs of selective reporting in the literature on the effects of climate change on plant-herbivore interactions.

Another motivation for the examination of potential selective reporting is the controversy concerning the scientific consensus on anthropological climate change between John Cook and colleagues on one side and Richard Tol on the other. Cook *et al.* (2013) collect almost 12,000 abstracts from peer-reviewed studies and conclude that 97% of those support the argument that climate change is human-made. Tol (2014) disagrees and has reservations to the way how Cook *et al.* (2013) select papers for their survey. Cook *et al.* (2014), in turn, disagree with the response of Tol (2014) and point out several alleged mistakes in Tol's arguments. From our perspective the main problem of the Cook *et al.* (2013) survey is that it does not mention nor correct for potential selective reporting. Given how widespread the file-drawer problem is in many fields, the fact that 97% studies report positive results does not necessarily translate into a 97% consensus of the scientific community that climate change is human-made. Because our prior about the sign of the relation between human activity and climate change is so strong, researchers may be less inclined to report neutral than large positive estimates of the relationship. It is perhaps a case in point that before being accepted by Energy Policy, Tol's comment was rejected by Environmental Research Letters, the outlet where Cook *et al.* (2013) was published.[1]

In contrast to most subjects of meta-analysis in economics, the social cost of carbon is not estimated in a regression network. Rather, the SCC is a result of a complex calibration exercise, and the uncertainty surrounding the estimates is usually determined via Monte Carlo simulations. Therefore by definition the literature lacks the usual suspects when it comes to potential selective reporting: specification search across models with different control variables, choice of the estimation technique, and the selection of the data sample. On the other hand, the authors have the liberty to choose among many possible values of the parameters that enter the computation and influence both the estimated magnitude of the SCC and the associated uncertainty. In a critical review of integrated assessment models, Pindyck (2013, p. 863) argues that "these models can be used to obtain almost any result one desires." Despite the difficulty in computing the SCC, we believe it is worth trying to pin down this crucial parameter. Testing for the potential selective reporting bias represents a part of this effort.

The remainder of the paper is structured as follows. Section 7.2 briefly discusses how the authors derive estimates of the social cost of carbon. Section 7.3 describes how we collect data for the meta-analysis. Section 7.4 explains the methods used in economics for the detection of selective reporting and addresses the specifics of their application in the case of the social cost of carbon. Section 7.5 presents the results of meta-regression analysis based on the tests of funnel asymmetry. Section 7.6 concludes the paper. A list of studies included in the meta-analysis and summary statistics of regression variables are reported in the Appendix.

[1] As Richard Tol describes on his website: http://richardtol.blogspot.com/2013/06/draft-comment-on-97-consensus-paper.html.

7.2 Estimating the Social Cost of Carbon

The purpose of this section is to outline the intuition behind the estimation of the SCC and discuss the results of the related literature, not to provide a detailed overview of estimation methodology. For the latter we refer the reader to Pindyck (2013) and Greenstone *et al.* (2013).

The first estimate of the shadow price of carbon emissions dates back to Nordhaus (1982). In the early 1990s William Nordhaus developed the first predecessor of the current generation of models, Nordhaus (1991), which he applied to the US economy. Later, Nordhaus extrapolated his country-level estimates of welfare effects to a global estimate, which has become the norm in the literature. Several researchers followed this approach (for example, Ayres & Walter 1991), but it was not before Fankhauser (1994) that an uncertainty component was introduced into the analysis. In the following years the literature differentiated further and more distinct models were introduced: among others, Tol (1995), Nordhaus & Yang (1996), and Plambeck & Hope (1996).

The workhorse tool for the estimation of the SCC are the so-called integrated assessment models. In simple terms, an integrated assessment model puts the expected climate effects of carbon emissions into the framework of economic growth theory. The social cost of carbon is then calculated as the difference between the present and future GDP influenced by damages resulting from carbon emissions, discounted back to the present time. The three most commonly used models are DICE [Dynamic Integrated Climate and Economy] developed by William Nordhaus (Nordhaus 2008), PAGE [Policy Analysis of the Greenhouse Effect] developed by Chris Hope (Hope 2008b), and FUND [Climate Framework for Uncertainty, Negotiation, and Distribution] developed by Richard Tol (Tol 2002a;b). Each model specifies how climate impacts result in economic damages in a different way (for more details on the differences in methodology see, for example, NRC 2009; IWG 2010; 2013).

The mapping of carbon emissions to economic costs is associated with significant uncertainties. The authors must rely on trends and scenarios taken from other sources, which involves simplification of complex processes. The authors must make assumptions about the level of current and future emissions (under different scenarios), about how these emissions translate into atmospheric gas concentrations (resulting from current, past, and future emissions), how these concentrations translate into warming (climate sensitivity), and how the warming translates into economic damages (projections of technological change, social utility assumptions, and damage functions). A major source of uncertainty is linked to the discount rate in monetary valuations. The resulting SCC is either a best-guess value of the calibration provided by the researcher or a mean/median value with a probability distribution, usually constructed using a Monte Carlo simulation. The reported values of the SCC vary widely.

Several attempts have been made to synthesize the published information on the optimal carbon tax. The IPCC (1995) literature review reports the range of best guesses from existing studies published until 1995: for carbon emitted in 1995, the range of estimates covers 5–125 USD/tC (in 1990 prices). In IPCC (2007), the values for 2005 emissions are extracted from about 100 estimates and range from -11 USD/tC to 348 USD/tC with an average value of 44 USD/tC (in 2005 prices). Both studies find the net damage costs of climate change to be significant and increasing over time. The IPCC emphasizes that these intervals do not represent the full range of uncertainty.[2]

The first comprehensive meta-analysis on the topic, Tol (2005b), collects 103 estimates from 28 different studies. Combining all the estimates into a composite probability density function, Tol (2005b) finds a median estimate of 14 and mean of 93, not exceeding 350 with a 95% probability. The estimates are driven by the choice of the discount rate and equity weights; Tol (2005b) also finds that the largest estimates with substantial uncertainty come from studies not published in peer-reviewed journals. In an update of the meta-analysis, Tol (2008) confirms his previous findings using 211 estimates collected from 47 studies; moreover, he identifies a downward trend in the reported SCC. Using the Fisher-Tippett fat-tailed distribution for the

[2]The fifth assessment report, IPCC (2014), refers to the updated meta-analyses by Richard Tol.

probability density function, for emission year 1995 discounted to 1995 he estimates the median SCC at 74 and the mean at 127, not exceeding 453 with a probability of 95%.

In another update, Tol (2011) performs a meta-regression analysis of 311 estimates of the social cost of carbon. He estimates a global mean SCC to be 177 (in 2010 USD and for 2010 emission year), median to be 116 with a standard deviation of 293, not exceeding 669 USD/tC with a 95% probability. A lower discount rate leads to a higher social cost of carbon, and peer-reviewed estimates and estimates from newer studies seem to be less pessimistic. In the most recent survey, Tol (2013b) adds another 277 estimates from 14 studies to the meta-analysis and gets a mean estimate of 196 and a median of 135 with a standard deviation of 322.

7.3 The SCC Data Set

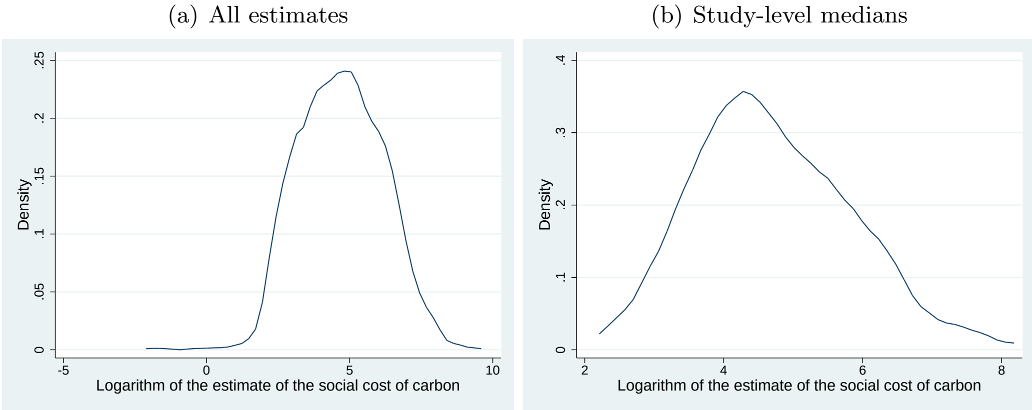
The first step of any meta-analysis is the collection of results from primary studies that report estimates of the effect in question. We take the advantage of the previous meta-analyses of the literature estimating the social cost of carbon and start with the data set provided by Richard Tol. The data set covers studies published until mid-2012 and includes 79 papers. Additionally we search in Google Scholar for new studies published in 2012 and later; the search query is available in the online appendix. We identify 22 new studies, bringing the total number of papers included in the meta-analysis to 101, listed in the Appendix. Most studies report multiple estimates of the social cost of carbon; for example, with different assumptions concerning the pure rate of time preference or different economic scenarios. We collect all of the estimates, which yields 809 observations. To put these numbers into perspective, we refer to the recent survey of meta-analyses in economics, Doucouliagos & Stanley (2013), who note that the largest meta-analysis conducted so far uses 1,460 estimates from 124 studies.

Aside from collecting additional studies, we also make adjustments in the original data set provided by Tol. Some studies available as mimeographs at the time when Tol collected the data have been published since 2012, and for these studies we checked the reported results and, if needed, changed the coding of the data accordingly. We also collect additional variables that may help explain the heterogeneity in the estimates of the social cost of carbon. Because the estimates of the SCC are reported for different emission years and evaluated in nominal US dollars, we have to recompute them to a common metric. We choose 2010 as the price year and 2015 as the emission year; for the normalization of the emission year we assume a constant growth of the SCC of 3.11% per year, the mean growth of the estimated real SCC between emission years in our data set (more details are available in the online appendix). Some studies report the SCC as the cost of emission of a molecule of carbon dioxide, while others refer to the cost of emission of an atom of carbon. We recompute the estimates so that they relate to the cost per metric ton of carbon.

We add the last study to our data set on August 1, 2014. At that time all studies taken together had obtained almost 17 thousand citations in Google Scholar (or almost 1,700 on average per year), which shows the scientific impact of the literature estimating the SCC. The first estimate was reported in 1982, but the median study in our data set comes from 2008: more and more studies on the topic are reported each year. Out of the 101 studies in our sample, 63 are published in peer-reviewed journals; the remaining 38 studies are book chapters, government reports, mimeographs, and other publications for which peer review is not guaranteed. We include the latter group of studies as well, partly following the advice of Tom Stanley to "better err on the side of inclusion in meta-analysis" (Stanley 2001, p. 135) and partly because we are interested in any potential differences in selective reporting between published and unpublished studies. Our approach to data collection and analysis is consistent with the Meta-Analysis of Economics Research Reporting Guidelines (Stanley *et al.* 2013).

Figure 7.1 shows the distribution of the estimates of the social cost of carbon in our data set. Because the distribution is skewed to the right (the mean estimate is 290, the median is 99), we choose the logarithmic scale for the depiction of the data set. To be able to take the log of all estimates, we add 13 to the observations

Figure 7.1: Kernel density plots



Notes: Because the smallest estimate in our data set is −12.8, we add 13 to all estimates of the social cost of carbon before taking logs.

(the smallest estimate is −12.8). Panel A of Figure 7.1 shows the distribution for all estimates; Panel B shows the distribution of study-level medians reported in studies: both distributions are approximately log-normal, which is corroborated by the skewness and kurtosis test of normality, although the distribution of medians is slightly skewed to the right even after taking logs. The mean and median of study-level median estimates are smaller than those of all estimates (201 vs. 290 and 82 vs. 99, respectively), which suggests that studies which obtain larger SCC in general report more estimates.

Figure 7.2 depicts the box plot of the estimates of the SCC reported in individual studies. Even with the logarithmic scale, the figure shows substantial heterogeneity across studies. It follows that it is important to control for the methodology of the SCC computation employed in the study and to cluster standard errors in the resulting regressions at the study level, because estimates reported within individual studies are unlikely to be independent. All variables that we collect for this meta-analysis are summarized and explained in Table 7.1; the table corresponds to the entire data set of 809 observations. Summary statistics for the two additional data sets (study medians and estimates with reported uncertainty) are shown in the Appendix.

Figure 7.2: Estimates of the social cost of carbon vary

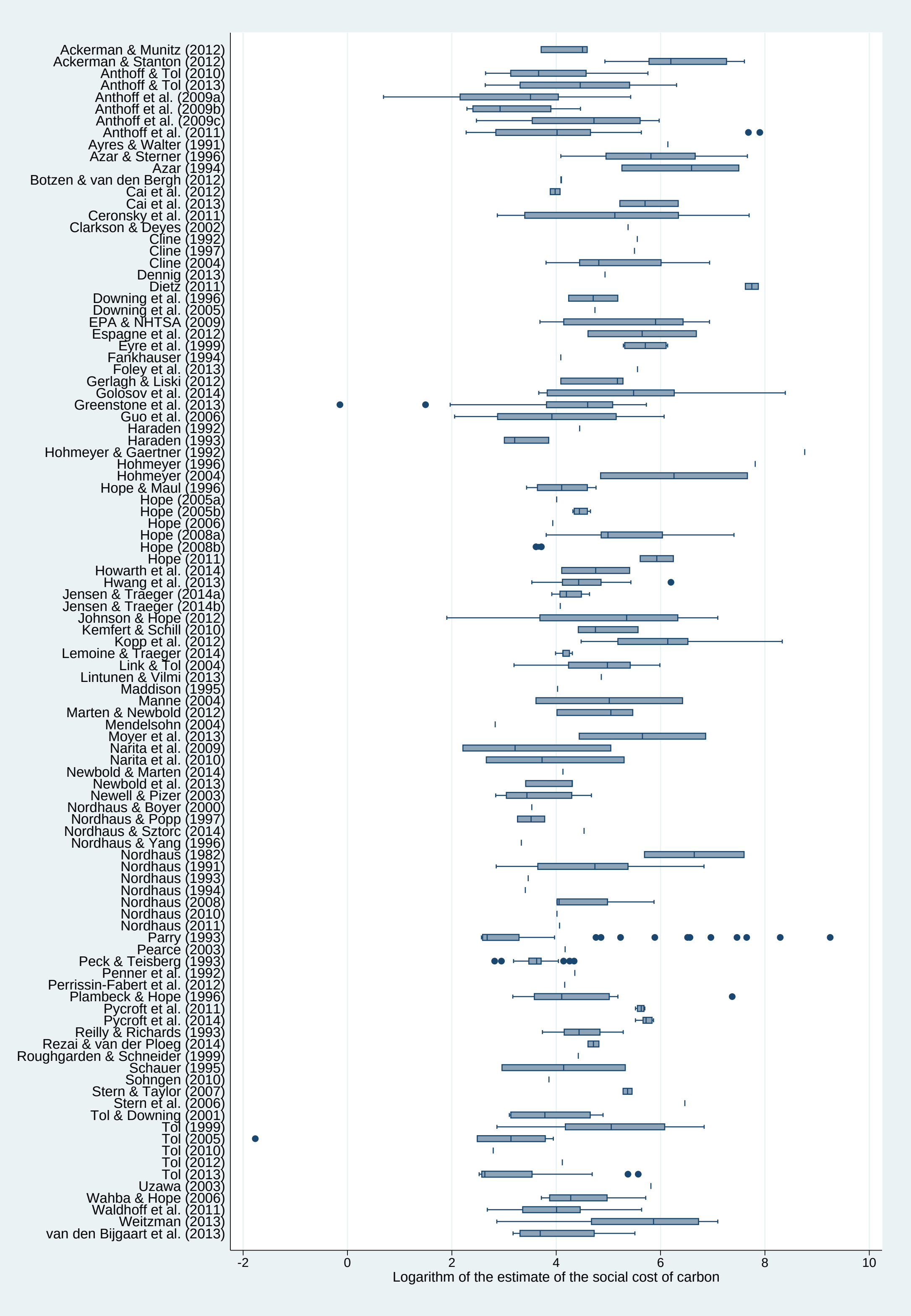


Table 7.1: Description and summary statistics of regression variables

Variable	Description	Obs.	Mean	Std. dev.
SCC	The reported estimate of the social cost of carbon in USD per ton of carbon (normalized to 2015 emission year in 2010 dollars).	809	290	635
Standard error	The approximate standard error of the estimate computed from the reported lower bound of the confidence interval.	267	162	235
Upper SE	The approximate standard error of the estimate computed from the reported upper bound.	267	1182	1921
Reviewed	= 1 if the study was published in a peer-reviewed outlet.	809	0.80	0.40
Publication year	The year of publication of the study (base: 1982).	809	24.7	7.46
Mean estimate	= 1 if the reported SCC estimate is the mean of the distribution.	809	0.23	0.42
Median estimate	= 1 if the reported SCC estimate is the median of the distribution.	809	0.21	0.41
Marginal costs	= 1 if the study estimates marginal damage costs (damage from an additional ton of carbon emitted) rather than average costs (the total impact divided by the total emissions of carbon).	809	0.96	0.20
Dynamic impacts	= 1 if the study examines dynamic impacts of climate change or uses a dynamic model of vulnerability.	809	0.40	0.49
Scenarios	= 1 if the study uses climate and economic scenarios that are internally consistent. A few studies use arbitrary assumptions about climate change.	809	0.82	0.39
FUND	= 1 if the authors use the FUND model or derive their model from FUND.	809	0.40	0.49
DICE or RICE	= 1 if the authors use the DICE/RICE model or derive their model from DICE/RICE.	809	0.46	0.50
PAGE	= 1 if the authors use the PAGE model or derive their model from PAGE.	809	0.19	0.39
PRTP	The pure rate of time preference assumed in the estimation.	633	1.23	1.57
Equity weights	= 1 if equity weighting is applied.	809	0.18	0.38
Pigovian tax	= 1 if the estimate is computed along a trajectory of emissions in which the marginal costs of emission reduction equal the SCC, then the estimate corresponds to a Pigovian tax.	809	0.29	0.45
Citations	= The logarithm of the number of Google Scholar citations of the study.	809	3.54	1.30
Journal rank	= SciMago journal rank based on the impact factor extracted from Scopus.	809	1.32	2.33

Notes: Data are collected from studies estimating the social cost of carbon. The data set is available at `meta-analysis.cz/scc`.

The construction of the approximate standard errors for the estimates of the social cost of carbon (the second and third item in Table 7.1) will be described in detail in the following two sections. We can only approximate standard errors for estimates for which the authors of primary studies report a measure of uncertainty, usually a confidence interval. Only 267 out of 809 estimates in our data set are reported together with a measure of uncertainty. These estimates are on average much larger than the rest of the data: the mean estimate with uncertainty is 411 (in contrast with 290 when all estimates are considered) and the median is 241 (in contrast with 99). In other words, authors who provide a probabilistic distribution of estimates tend to report much larger median values of the SCC than authors who only report their best-guess estimates.

We include a dummy variable to take into account whether the study in which the estimate is reported is published in a peer-reviewed journal. We also control for the year of publication of the study: perhaps novel methods of estimating the SCC bring systematically different results, and the literature converges to a consensus value. We include dummy variables for the case when the reported estimate corresponds to the median and mean of the distribution; the base category corresponds to best-guess estimates. Some studies estimate average costs rather than marginal damage costs, and we control for this aspect of methodology as well. We include dummy variables for studies that examine dynamic impacts of climate change and studies that use internally consistent climate and economic scenarios to simulate the evolution of emissions.

Three families of integrated assessment models are predominant in the estimation of the social cost of carbon: the FUND, PAGE, and DICE (RICE) models; most author teams also use consistently the same family of models. We include three dummy variables to distinguish between these approaches. Some estimates are constructed as weighted averages of several model approaches, and a few studies use models independent of the main three families. An important feature in estimating the SCC is the assumed discount rate, especially the rate of pure rate of time preference—we control for the value assumed in the computation, but some authors do not report it; we only have data on the pure rate of time preference for 633 estimates. Next, some studies employ equity weights in the computation, and we control for this aspect of methodology. We also include a dummy variable that equals one if the estimate corresponds to the optimal abatement path and can be interpreted as the Pigovian tax on carbon emissions. Finally, we control for the number of Google Scholar citations of the study and the SciMago journal rank of the outlet (the SciMago journal rank based on Scopus citations is available for more journals in our sample than the Thompson Reuters impact factor and the RePEc impact factor): perhaps these study characteristics capture aspects of quality not covered by the methodology variables introduced above.

In the next step we examine how method and publication characteristics are correlated with the reported estimates of the SCC. The first two columns of Table 7.2 report the results of a regression of the estimates on the estimates' characteristics; the third and fourth column use a logarithm of the estimate of the SCC on the left-hand side of the regression. In all cases we cluster standard errors at the study level to take into account within-study correlation in SCC estimates. The results suggest that studies published in peer-reviewed journals report, on average, substantially smaller estimates of the social cost of carbon. This evidence is consistent with previous research (Tol 2011), and can be interpreted in two ways. The first potential interpretation, suggested by Tol (2011), puts forward that many large estimates of the SCC that we observe in the literature are not verified by the peer-review process, and thus may be of questionable quality. The second possible interpretation, in line with the topic of this paper, would suggest that the peer-review process results in a selective reporting bias in favor of more conservative estimates of the SCC. We will examine this issue in detail in the next two sections.

Table 7.2 also shows that the year of publication is not systematically related to the magnitude of the reported SCC. (We also experimented with several specifications that were nonlinear in the year of publication, but obtained no statistically significant results.) In contrast, Tol (2011) finds that newer studies tend to report smaller estimates of the SCC. Our results are different because we include new studies published between 2012 and 2014; these studies often report large estimates of the SCC as they try to incorporate potential catastrophic outcomes of climate change. Next, we find that authors who report uncertainty associated with their central estimates (usually confidence intervals around mean or median expected SCC values) tend to report larger SCC. The evidence on the importance of estimating marginal instead

Table 7.2: Explaining the heterogeneity in the SCC estimates

	SCC		log SCC	
	All estimates	PRTP	All estimates	PRTP
Reviewed	-187.1***	-149.2*	-0.741***	-0.574**
	(65.34)	(78.37)	(0.225)	(0.253)
Publication year	-4.877	-4.004	0.0212	0.0241
	(6.595)	(7.129)	(0.0177)	(0.0246)
Mean estimate	138.8***	256.7***	0.439**	0.914***
	(52.64)	(65.96)	(0.182)	(0.227)
Median estimate	316.4***	243.0***	1.366***	1.185***
	(76.60)	(72.92)	(0.252)	(0.306)
Marginal costs	-331.7	-380.7	-1.204***	-1.179***
	(272.0)	(287.2)	(0.387)	(0.414)
Dynamic impacts	-213.1***	-330.0**	-0.482*	-0.946**
	(78.70)	(152.5)	(0.272)	(0.429)
Scenarios	140.5	199.8	0.745***	0.676*
	(124.3)	(148.2)	(0.235)	(0.357)
FUND	45.66	33.65	-0.270	-0.202
	(99.22)	(140.0)	(0.295)	(0.393)
DICE or RICE	75.01	-70.24	0.240	-0.531
	(56.30)	(84.98)	(0.160)	(0.340)
PAGE	-173.2**	-304.9**	-0.147	-0.679*
	(76.14)	(145.7)	(0.199)	(0.353)
Equity weights	31.31	73.26	0.392*	0.554**
	(52.89)	(71.41)	(0.202)	(0.262)
Pigovian tax	-85.01	-46.26	-0.226	0.137
	(81.76)	(72.78)	(0.253)	(0.295)
Citations	-20.58	-24.49	0.0568	0.116
	(29.55)	(32.32)	(0.0775)	(0.0790)
Journal rank	36.43***	26.02*	0.102***	0.0107
	(8.943)	(13.98)	(0.0270)	(0.0402)
PRTP		-112.7***		-0.425***
		(22.64)		(0.0913)
Constant	774.6**	999.1**	4.800***	5.384***
	(366.4)	(431.6)	(0.633)	(0.695)
Observations	809	633	809	633

Notes: The table presents the results of regression $SCC_{ij} = \alpha + \beta \cdot X_{ij} + u_{ij}$, where SCC_{ij} is the i-th estimate of the social cost of carbon reported in the j-th study and X is a vector of the estimate's characteristics. In the last two columns we use the logarithm of the estimates of SCC as the dependent variable; because the smallest estimate in our data set is -12.8, we add 13 to all estimates of the social cost of carbon before taking logs. Estimated by OLS; standard errors are clustered at the study level and shown in parentheses. PRTP = only estimates for which the authors report the pure rate of time preference used in the computation. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

of average costs is mixed: we only find significant results in the case of log-level regressions, which suggest that estimating average costs exaggerates the reported SCC. Authors investigating dynamic impacts of climate change report, on average, smaller estimates of the SCC.

Studies employing internally consistent economic and climate scenarios tend to report larger estimates of the SCC, but the effect is only statistically significant in the log-level specifications of the regression. There is also some evidence that authors employing a variant of the PAGE model report, ceteris paribus, smaller estimates of the SCC than other studies, but the effect is not statistically significant at the 5% level in all specifications. The log-level regressions suggest that using equity weights results in larger reported SCC. In contrast, it does not seem to be important for the magnitude of the estimated SCC whether the estimate is consistent with the optimal abatement path and thus represents a Pigovian tax. Similarly the number of citations of the study is not systematically related to the reported results. The ranking of the journal, on the other hand, is correlated with the estimated SCC: studies published in better journals tend to report larger estimates. Finally, as expected, a larger assumed pure rate of time preference leads to smaller estimates of the SCC.

7.4 Detecting Selective Reporting

In this section we overview the tools that are available for the examination of selective reporting in economics. Three methods are commonly used to detect potential selective reporting bias in the literature: Hedges' model, the funnel plot, and meta-regression analysis. Concerning the first method, Hedges (1992) introduces a model of selective reporting which assumes that the probability of reporting of estimates is determined by their statistical significance. The probability of reporting only changes when a psychologically important p-value is reached: in economics these threshold values are commonly assumed to be 0.01, 0.05, and 0.1. When no reporting bias is present, all estimates, significant and insignificant at conventional levels, should have the same probability of being published. The augmented model developed by

Ashenfelter *et al.* (1999) allows for heterogeneity in the estimates of the underlying effect. The augmented log-likelihood function is (Ashenfelter *et al.* 1999, p. 468):

$$L = c + \sum_{i=1}^{n} \log w_i(X_i, \omega) - \frac{1}{2} \sum_{i=1}^{n} \Big(\frac{X_i - \mathbf{Z}_i\Delta}{\eta_i}\Big)^2 - \sum_{i=1}^{n} \log(\eta_i) - \sum_{i=1}^{n} \log \Big[\sum_{j=1}^{4} \omega_j B_{ij}(\mathbf{Z}_i\Delta, \sigma)\Big], \tag{7.1}$$

where $X_i \sim N(\Delta, \eta_i)$ would be the estimates of the social cost of carbon. The parameter Δ is the average underlying SCC, and $\eta_i = \sigma_i^2 + \sigma^2$, where σ_i are the reported standard errors of the estimates and σ measures heterogeneity in the estimates. The probability of reporting is determined by the weight function $w(X_i)$. In this model $w(X_i)$ is a step function associated with the p-values of the estimates. $B_{ij}(\Delta, \sigma)$ represents the probability that an estimate X_i will be assigned weight ω_i. For the first step, p-value < 0.01, ω is normalized to 1 and the author evaluates whether the remaining three weights differ from this value. Z_i is a vector of the characteristics of estimate X_i. In the absence of selective reporting the meta-analyst is not able to reject the hypothesis $\omega_2 = \omega_3 = \omega_4 = 1$; that is, estimates with different levels of statistical significance have the some probability of being reported.

The second method of detecting selective reporting is a visual examination of the so-called funnel plot (Egger *et al.* 1997). The funnel plot is a scatter plot of the estimated coefficients (in our case the reported estimates of the social cost of carbon) on the horizontal axis and their precision (the inverse of standard error) on the vertical axis. The most precise estimates are close to the top of the funnel and are tightly distributed. As precision decreases, the dispersion of estimates increases, which yields the shape of an inverted funnel with a sharp tip on the top and a wide base on the bottom. In the absence of selective reporting the funnel should be symmetrical: all imprecise observations have the same probability of being reported. Even if the true effect is positive, due to the laws of chance we should observe some negative estimates with low precision (as well as large estimates with low precision). If, in contrast, some estimates (for example, the negative ones) are systematically omitted, the funnel becomes asymmetrical.

The third method used to investigate potential selective reporting is closely related to the funnel plot, but uses meta-regression analysis to statistically examine the degree of funnel asymmetry. When selective reporting is absent from the literature the estimates of the SCC will be randomly distributed around the true mean estimate of the social cost of carbon, SCC_0 (due to the central limit theorem). But if authors discard some estimates because they are statistically insignificant at the conventional levels or have a sign that is inconsistent with the theory or the mainstream prior, the reported estimates of the SCC will be correlated with their standard errors (Card & Krueger 1995):

$$SCC_i = SCC_0 + \beta_0 \cdot Se(SCC_i) + u_i, \tag{7.2}$$

where SCC_i is the estimate of the social cost of carbon, SCC_0 denotes the average underlying value of the social cost of carbon, $Se(SCC_i)$ denotes the standard error of SCC_i, β_0 measures the magnitude of selective reporting, and u_i is an error term. Specification (7.2) can be thought of as a test of the asymmetry of the funnel plot: the regression results from rotating the axes of the funnel plot and inverting the values on the new horizontal axis. A statistically significant estimate of β_0 provides formal evidence for funnel asymmetry, and thus for selective reporting. Note that β_0 close to 2 is consistent with a situation when only positive and statistically significant SCC estimates (that is, the estimates for which the corresponding 95% confidence intervals exclude zero) are selected for reporting and other estimates are hidden in file drawers. Since specification (7.2) is heteroscedastic (the dispersion of the dependent variable increases when the values of the independent variable increase), in practice meta-analysts often estimate it by weighted least squares with precision taken as the weight (Stanley 2005):

$$SCC_i/Se(SCC_i) = t_i = SCC_0 \cdot 1/Se(SCC_i) + \beta_0 + \xi_i. \tag{7.3}$$

Because most studies provide more than one estimate of the SCC, it is important to take into account that estimates reported in one study are likely to be correlated. A way how to address this issue is to employ the so-called mixed-effects multilevel model (for an early application in meta-analysis, see, for example, Doucouliagos &

Stanley 2009), which assumes unobserved between-study heterogeneity. We specify the mixed-effects model following Havranek & Irsova (2011) and Havranek *et al.* (2012):

$$t_{ij} = e_0 \cdot 1/Se(SCC_{ij}) + \beta_0 + \zeta_j + \epsilon_{ij}, \tag{7.4}$$

where i and j denote estimate and study subscripts and t_i denotes the approximate t-statistic. The overall error term (ξ_{ij}) now breaks down into study-level random effects (ζ_j) and estimate-level disturbances (ϵ_{ij}). The model is estimated by restricted maximum likelihood. The problem of the mixed-effects model is that it assumes no correlation between study-level random effects and the independent variables. This assumption is rarely tenable in practice, and we thus prefer to run the fixed-effects model and cluster standard errors at the study level.

The three methods of detecting selective reporting introduced above are designed for regression estimates of the parameter in question and require that the ratio of the point estimate to the standard error be t-distributed. In contrast, estimates of the social cost of carbon are based on calibration and assumptions concerning the uncertainty about parameters entering the computation. For most estimates of the SCC the authors do not report confidence intervals, and even if they do, we cannot assume the ratio of the point estimate to the standard error to have a t-distribution because of the asymmetries in uncertainty surrounding the SCC (especially catastrophic events). In particular, Hedges' method assumes that authors decide on which estimates to report depending on whether the estimates surpass a certain threshold of the p-value, which is unlikely to be the driving factor of selective reporting in the literature on the SCC. In contrast, we can use the intuition behind the two methods based on the analysis of funnel plot asymmetry.

To be able to employ the methods based on the funnel plot, we need to compute the approximate standard errors of the estimates. Few authors report the standard errors directly, and only 267 out of 809 estimates are reported together with a measure of uncertainty from which confidence intervals can be computed (usually 95% confidence intervals). The confidence intervals of the estimates of the SCC are typically asymmetrical, which means that for the approximation of the standard error we have to choose whether we will use the lower or upper bound of the confidence interval. We choose the lower bound, because we assume that any potential selective reporting in the literature will be associated with the sign of the estimate and the authors' confidence that the true SCC is nonzero.[3] We additionally examine whether the asymmetry of the confidence intervals reported by the authors affects our results concerning potential selective reporting in the literature; a similar problem in the analysis of selective reporting is also discussed in detail by Rusnak *et al.* (2013).

Because for most estimates of the SCC the authors do not report confidence intervals or other measures of uncertainty, we also choose an alternative approach for the computation of approximate standard errors. From each study we take the median estimate of the SCC and then construct the standard error as the difference between the 50^{th} and the 16^{th} percentile of the distribution of estimates. (We only use studies that report multiple estimates of the SCC.) The standard errors are computed under the simplifying assumption that estimates in each study are normally distributed. Most studies produce an asymmetric distribution of estimates, but we are interested in quantifying the confidence of the authors that their estimate of the social cost of carbon is different from zero, which is analogous to statistical significance for classical regression estimates used in economic meta-analyses. We expect that selective reporting in the literature would manifest itself as a tendency to report less uncertainty (a smaller approximate standard error computed from the lower bound of the confidence interval or the distribution of estimates in a study) for smaller estimates of the SCC in the absolute value. As far as we know, this paper represents the first attempt to quantify potential selective reporting among simulated model results.

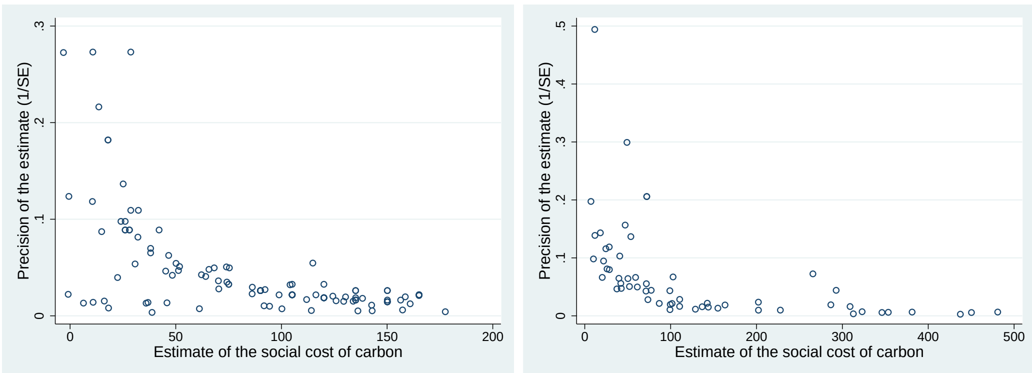
[3] Note that the social costs of carbon may, in principle, be negative. If the adverse consequences of climate change are small enough, they are offset by boosted yields in agriculture generated by increased atmospheric concentration of carbon dioxide. Several studies produce negative estimates of the SCC in some scenarios; for example, Tol (2005a); Anthoff *et al.* (2009a); Greenstone *et al.* (2013).

7.5 Meta-Regression Results

Figure 7.3 reports two funnel plots for the literature estimating the social cost of carbon: the funnel in panel A corresponds to estimates for which the authors report a measure of uncertainty, the funnel in panel B corresponds to study-level medians computed from all observations reported in the study. Both scatter plots resemble a right-hand part of an inverted funnel; the left-hand part is missing: few negative estimates of the social cost of carbon are reported. The funnels are clearly asymmetrical, with smaller estimates being typically more precise—that is, reporting less uncertainty in the downward direction. Large point estimates of the SCC are usually associated with a lot of uncertainty and do not exclude the possibility of small positive SCC. It is remarkable that the funnels have a similar shape even though the method of computing approximate standard errors differs a lot for the two cases.

Figure 7.3: Funnel plots show signs of selective reporting

(a) Estimates with uncertainty (b) Study-level medians



Notes: In the absence of selective reporting the funnel should be symmetrical around the most precise estimates of the social cost of carbon. Precision is the inverse of the approximate standard error computed from the lower bound of the reported confidence interval (or from the distribution of estimates in the study in the case of study-level medians). Outliers are excluded from the figure but included in all statistical tests.

Panel A of Table 7.3 shows the results of funnel asymmetry tests for the sample of estimates with uncertainty; in all specifications we cluster standard errors at the study level. In the first column we run a simple OLS regression of point estimates of the SCC on the approximate standard errors. The slope coefficient in the regression is positive and statistically significant, which corroborates our intuition based on funnel plots: larger estimates of the SCC are associated with larger downward uncertainty, and vice versa. The estimated slope coefficient equals approximately 1.7, which corresponds to "substantial" selective reporting bias according to the classification by Doucouliagos & Stanley (2013). We have noted that the slope coefficient close to 2 would be consistent with a situation when researchers systematically omitted estimates for which the 95% confidence interval included zero.

The constant in the regression can be interpreted as the mean estimate of the SCC when uncertainty about the SCC approaches zero (that is, corrected for any potential selective reporting), and is large and statistically significant in this specification, though smaller than the simple mean of all estimates. In the second column we add study-level fixed effects; in this way we filter out all study-specific characteristics that may influence the reported estimates. The result concerning the extent of selective reporting is similar to the previous case, but the estimate of the underlying SCC is now statistically insignificant at conventional levels.

Table 7.3: Funnel asymmetry tests, estimates with uncertainty

Panel A	OLS	FE	Precision	Study	ME
Standard error	1.705**	1.889**	2.467***	1.213**	1.819***
	(0.630)	(0.762)	(0.480)	(0.527)	(0.0825)
Constant	134.1**	104.2	10.27	63.14	-18.69
	(58.16)	(123.9)	(7.361)	(40.12)	(48.43)
Observations	267	267	267	267	267
Panel B	OLS	FE	Precision	Study	ME
Standard error	1.662**	1.907**	2.451***	0.780	1.835***
	(0.663)	(0.779)	(0.538)	(0.548)	(0.0843)
Upper SE	0.0246	-0.0109	0.00283	0.222	-0.00788
	(0.0254)	(0.00676)	(0.0107)	(0.143)	(0.0100)
Constant	112.0**	114.1	9.555	45.29	-17.78
	(50.00)	(118.6)	(6.133)	(29.63)	(48.81)
Observations	267	267	267	267	267

Notes: Panel A presents the results of regression $SCC_{ij} = SCC_0 + \beta \cdot SE(SCC_{ij}) + u_{ij}$, where SCC_{ij} is the i-th estimate of the social cost of carbon reported in the j-th study and $SE(SCC_{ij})$ is the corresponding approximate standard error computed from the lower bound of the reported confidence interval. Panel B presents the results of regression $SCC_{ij} = SCC_0 + \beta \cdot SE(SCC_{ij}) + \gamma \cdot SE^{up}(SCC_{ij}) + u_{ij}$, where $SE^{up}(SCC_{ij})$ is the corresponding approximate standard error computed from the upper bound of the reported confidence interval. The standard errors of regression parameters are clustered at the study level and shown in parentheses. FE = study level fixed effects. Precision = weighted by the inverse of the standard error. Study = weighted by the inverse of the number of estimates reported per study. ME = study-level mixed effects. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

In the next specification we weight estimates by their precision—the inverse of the approximate standard error. This weighted-least-squares specification has two benefits, for which it has commonly been used in meta-analysis: see, for example, Stanley (2005). First, it corrects for heteroskedasticity in the baseline regression, where the independent variable (the standard error of the estimate of the SCC) is a measure of dispersion of the dependent variable (the magnitude of the estimate of the SCC). Second, by definition it gives more weight to more precise results, which further alleviates the effects of selective reporting. The results are similar to the previous specification, but the coefficient associated with selective reporting is even larger—2.5, which corresponds to "severe" selective reporting based on the guidelines by Doucouliagos & Stanley (2013)—and statistically significant at the 1% level.

In the fourth column we use weighted least squares again, but instead of precision the weight is now the inverse of the number of estimates reported in each study. In unweighted regressions, studies that report many estimates get overrepresented and influence the results more heavily than studies with few reported estimates. Weighting by the inverse of the number of estimates reported per study seems natural because it gives each study approximately the same influence on the results. Compared to the baseline OLS regression, this specification yields smaller estimates of both the selective reporting parameter and the underlying mean SCC. The coefficient representing selective reporting is still statistically significant at the 5% level, and its extent would still be classified as substantial. In contrast, the coefficient that captures the mean effect corrected for the selective reporting bias is not statistically significant at conventional levels.

Finally we also employ the mixed-effects multilevel model and report the results in the last column of panel A in Table 7.3 . The mixed-effects model allows for random differences in the extent of the underlying SCC across studies and also gives each study approximately the same weight. The results corroborate the evidence reported in the previous columns concerning statistically significant and substantial selective reporting. The estimate of the underlying value of the social cost of carbon is once again statistically insignificant, and here even negative.

In panel B of Table 7.3 we examine whether our results concerning selective reporting are influenced by the asymmetry of confidence intervals that the authors report for their estimates of the social cost of carbon. The asymmetry of confidence intervals reported in individual studies is not an issue per se: many applications of meta-analysis quote the central limit theorem, which would imply that estimates should be symmetrically distributed in the absence of selective reporting even if the individual distributions were skewed. The problem is that the crucial assumption of the central limit theorem, independence of individual studies and estimates, is unlikely to hold in this case.

Table 7.4: Funnel asymmetry tests, study-level medians

	OLS	Precision	OLS	Precision
Standard error	1.506***	1.936***	1.502***	1.958***
	(0.372)	(0.307)	(0.413)	(0.307)
Upper SE			0.00387	-0.0295***
			(0.0496)	(0.00540)
Constant	61.07***	21.06***	60.53***	26.01***
	(16.47)	(5.957)	(15.28)	(6.069)
Observations	68	68	68	68

Notes: Columns 1 and 2 present the results of regression $SCC_j = SCC_0 + \beta \cdot SE(SCC_j) + u_j$, where SCC_j is the median estimate of the social cost of carbon reported in the j-th study and $SE(SCC_j)$ is the corresponding approximate standard error computed from the distribution of estimates in the study. Columns 3 and 4 present the results of regression $SCC_j = SCC_0 + \beta \cdot SE(SCC_j) + \gamma \cdot SE^{up}(SCC_j) + u_j$, where $SE^{up}(SCC_j)$ is the corresponding approximate standard error computed from the 84$^{\text{th}}$ percentile of the distribution of estimates in the study. The standard errors of regression parameters are robust to heteroskedasticity and shown in parentheses. Precision = weighted by the inverse of the standard error. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

To see whether asymmetry drives our results, we need to include an interaction term of the approximate standard error computed based on the lower bound of the confidence interval and the ratio of the standard error computed from the upper bound and from the lower bound. This means that we can simply add an independent variable that captures the approximate standard error computed based on the upper bound ($SE \cdot SE^{up}/SE = SE^{up}$), and Table 7.3 shows that it is statistically insignificant in all cases. All other results are qualitatively similar to the baseline regression, except for the specification where we use the inverse of the number of estimates reported per study as the weight—the coefficient corresponding to selective reporting loses statistical significance. In general, however, the results show that the evidence for selective reporting identified in the previous regressions is not substantially affected by the asymmetry of individual confidence intervals.

In Table 7.4 we repeat the previous exercise for study-level median estimates. In this setting, however, we have to omit the fixed-effects model, the mixed-effects model, and the weighted-least-squares regression with the inverse of the number of observations reported per study taken as the weight. Therefore we only report two sets of results, an OLS regression and a specification where estimates are weighted by their precision; both are run for the baseline relation between the estimates of the SCC and their standard errors and for the extended specification that includes the interaction of the standard error and the ratio of the upper and lower standard error (which simplifies to the upper standard error). The results concerning selective reporting are consistent with the evidence reported in Table 7.3: we obtain estimates of the selective reporting bias that are both statistically significant at the 5% and "substantial" according to the classification by Doucouliagos & Stanley (2013). In contrast to Table 7.3, however, we find consistently significant estimates of the mean SCC corrected for selective reporting: approximately between 20 and 60.

In Table 7.5 and Table 7.6 we examine whether our estimates of the magnitude of the selective reporting bias in the literature change when we control for additional aspects of estimates and studies. Table 7.5 focuses on the estimates for which the authors report a measure of uncertainty. In this setting we cannot use the fixed-effects specification, because some of the explanatory variables have the same value for all estimates reported in one study, so the variables would be perfectly correlated with individual study dummies. Note also that it makes little sense to interpret the constant in this regression; it still represents the mean value of the SCC corrected for selective reporting, but it is conditional on the values of all the other independent variables included in the regression. It is important that the estimates of the coefficient capturing selective reporting are consistent with the evidence reported in the previous tables: the estimates are statistically significant at the 5% level and lie in the range 1.2–2.3. The same findings hold in Table 7.6, where we use study-level

Table 7.5: Controlling for heterogeneity, estimates with uncertainty

	OLS	PRTP	Precision	Study	ME
Standard error	1.800***	1.899**	2.344***	1.227***	1.800***
	(0.628)	(0.731)	(0.534)	(0.439)	(0.0806)
Reviewed	195.6	193.2	48.76	-52.38	195.6*
	(123.8)	(135.1)	(42.35)	(125.5)	(111.2)
Publication year	-12.16	-15.47	-2.430	12.09	-12.16
	(18.66)	(20.65)	(2.341)	(13.23)	(8.480)
Mean estimate	350.1**	-373.3	33.29	-24.50	350.1**
	(157.0)	(309.8)	(31.73)	(131.4)	(137.3)
Median estimate	288.9*	-153.5	46.00*	-24.53	288.9**
	(145.8)	(238.8)	(26.16)	(105.1)	(131.2)
Marginal costs	-823.3*	-1041.4**	-64.37	-123.6	-823.3**
	(476.7)	(476.4)	(82.34)	(228.1)	(357.1)
Dynamic impacts	-303.7	-41.23	-101.7	-162.0	-303.7**
	(189.0)	(220.5)	(91.32)	(130.1)	(150.3)
Scenarios	411.7*	296.2***	31.09	387.2	411.7***
	(231.8)	(93.62)	(32.69)	(247.5)	(121.2)
FUND	202.8	753.3***	49.34	-1.745	202.8
	(144.7)	(209.3)	(95.21)	(138.2)	(160.7)
DICE or RICE	40.25	785.3*	-33.27	-112.8	40.25
	(114.9)	(402.9)	(30.39)	(123.6)	(99.38)
PAGE	-13.54	879.8**	-38.93	59.47	-13.54
	(100.4)	(399.9)	(28.10)	(77.51)	(83.10)
Equity weights	118.4	-50.70	17.53	-24.11	118.4
	(127.0)	(105.5)	(14.33)	(94.67)	(78.02)
Pigovian tax	213.2	-18.85	42.28	30.85	213.2**
	(148.6)	(61.46)	(36.31)	(100.5)	(95.60)
Citations	2.556	-65.95	-4.060	59.93	2.556
	(53.01)	(66.05)	(13.17)	(52.61)	(35.18)
Journal rank	-21.89	-6.780	-10.89	50.11	-21.89
	(50.63)	(67.52)	(10.81)	(70.51)	(45.80)
PRTP		-47.21			
		(35.44)			
Constant	255.3	868.6	79.47	-611.6	255.3
	(701.8)	(722.9)	(117.9)	(577.8)	(460.7)
Observations	267	217	267	267	267

Notes: The table presents the results of regression $SCC_{ij} = SCC_0 + \beta \cdot SE(SCC_{ij}) + \delta \cdot X_{ij} + u_{ij}$, where SCC_{ij} is the i-th estimate of the social cost of carbon reported in the j-th study, $SE(SCC_{ij})$ is the corresponding approximate standard error computed from the lower bound of the reported confidence interval, and X is a vector of the estimate's characteristics. Standard errors are clustered at the study level and shown in parentheses. OLS = an ordinary least squares regression using all estimates. PRTP = only estimates for which the authors report the pure rate of time preference used in the computation. Precision = weighted by the inverse of the standard error. Study = weighted by the inverse of the number of estimates reported per study. ME = study-level mixed effects. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

Table 7.6: Controlling for heterogeneity, study-level medians

	All estimates		PRTP	
	OLS	Precision	OLS	Precision
Standard error	1.589***	1.851***	1.654***	1.851***
	(0.425)	(0.375)	(0.495)	(0.446)
Reviewed	81.20	-16.86	93.95	-24.14**
	(83.47)	(13.09)	(71.90)	(9.920)
Publication year	10.73	0.764	11.98	1.031
	(7.146)	(0.607)	(9.059)	(0.742)
Mean estimate	16.47	-22.67	4.793	-9.382
	(28.62)	(21.41)	(56.41)	(19.02)
Median estimate	27.90	48.09	84.64*	3.734
	(38.79)	(46.46)	(49.34)	(26.10)
Marginal costs	-133.3	-26.95*	-160.0	-6.354
	(86.94)	(14.95)	(124.3)	(12.76)
Dynamic impacts	17.84	9.220	-58.19	-18.98
	(46.98)	(23.39)	(85.12)	(24.30)
Scenarios	6.820	28.19*	-62.67	-2.849
	(33.18)	(16.14)	(61.30)	(14.83)
FUND	-68.63	-27.48	104.7	6.847
	(56.58)	(31.29)	(75.86)	(23.96)
DICE or RICE	-45.27	29.69**	-7.099	10.32
	(66.20)	(14.09)	(67.90)	(14.38)
PAGE	136.4	44.90*	251.8	26.45
	(98.68)	(26.30)	(229.1)	(28.57)
Equity weights	-23.66	29.96	-64.86	13.23
	(82.51)	(19.56)	(116.8)	(16.38)
Pigovian tax	7.854	-13.88	54.04	-0.107
	(32.06)	(15.28)	(48.21)	(16.74)
Citations	34.00	-1.969	47.56	0.835
	(25.44)	(3.763)	(35.67)	(2.428)
Journal rank	-10.61	6.241*	-22.29	3.532
	(12.20)	(3.638)	(14.59)	(3.488)
PRTP			-23.29	4.893
			(34.28)	(8.478)
Constant	-256.9	7.316	-273.3	3.199
	(283.6)	(21.95)	(364.5)	(25.15)
Observations	68	68	53	53

Notes: The table presents the results of regression $SCC_j = SCC_0 + \beta \cdot SE(SCC_j) + \delta \cdot X_j + u_j$, where SCC_j is the median estimate of the social cost of carbon reported in the j-th study, $SE(SCC_j)$ the corresponding approximate standard error computed from distribution of estimates in the study, and X is a vector of the estimate's characteristics. Standard errors are robust to heteroskedasticity and shown in parentheses. PRTP = only estimates for which the authors report the pure rate of time preference used in the computation. Precision = weighted by the inverse of the standard error. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

medians and construct medians for the independent variables that are not defined at the study level.

In Table 7.7 we investigate whether publication characteristics are associated with selective reporting. To this end we use the baseline specification of the funnel asymmetry test and include interactions of the standard error and the number of citations, a dummy variable that equals one if the study is published in a peer-reviewed journal, and ranking of the journal. The results are consistent both for the sub-sample of estimates with uncertainty and for median estimates taken from individual studies: studies published in peer-reviewed journals tend to suffer more from selective reporting than unpublished papers. The number of citations and journal rank, in contrast, do not systematically influence the magnitude of the selective reporting bias.

Table 7.7: What drives selective reporting?

	Estimates with uncertainty			Study-level medians	
	OLS	Precision	ME	OLS	Precision
Standard error	0.793*	0.650	0.891***	1.342***	1.692***
	(0.427)	(0.536)	(0.178)	(0.204)	(0.426)
SE · Reviewed	3.409***	2.548***	3.593***	2.581***	1.386**
	(0.862)	(0.645)	(0.252)	(0.833)	(0.555)
SE · Citations	-0.494	-0.127	-0.548***	-0.130	-0.0990
	(0.300)	(0.248)	(0.109)	(0.110)	(0.133)
SE · Journal rank	-0.368	-0.453*	-0.297***	-0.269**	-0.0974
	(0.248)	(0.250)	(0.0860)	(0.118)	(0.0590)
Constant	44.92**	12.48**	-15.34	31.15	19.29***
	(21.25)	(5.814)	(37.13)	(19.68)	(5.997)
Observations	267	267	267	68	68

Notes: Columns 1–3 present the results of regression $SCC_{ij} = SCC_0 + \beta \cdot SE(SCC_{ij}) + \epsilon \cdot X_{ij} \cdot SE(SCC_{ij}) + u_{ij}$, where SCC_{ij} is the i-th estimate of the social cost of carbon reported in the j-th study, $SE(SCC_{ij})$ is the corresponding approximate standard error computed from the lower bound of the reported confidence interval, and X is a vector of the estimate's characteristics. Columns 4 and 5 present the results of regression $SCC_j = SCC_0 + \beta \cdot SE(SCC_j) + \epsilon \cdot X_j \cdot SE(SCC_j) + u_j$, where SCC_j is the median estimate of the social cost of carbon reported in the j-th study, $SE(SCC_j)$ is the corresponding approximate standard error computed from the distribution of estimates in the study, and X is a vector of the estimate's characteristics. The standard errors of regression coefficients are clustered at the study level (or robust to heteroskedasticity in columns 4 and 5) and shown in parentheses. Precision = weighted by the inverse of the standard error. ME = study-level mixed effects. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

The finding that selective reporting is associated more with published studies than unpublished manuscripts could indicate that self-censorship is not the only source of selection in the literature on the social cost of carbon. The results are consistent with a situation when journal editors or referees prefer estimates of the SCC that are conclusive; that is, the estimates for which the approximate 95% confidence interval excludes zero. Nevertheless, the same pattern would be achieved through self-censorship if the authors believed that editors and referees preferred conclusive estimates and, therefore, selected such estimates for submission to journals.

7.6 Concluding Remarks

In this paper we conduct a meta-analysis of the literature estimating the social cost of carbon. We examine 809 estimates of the SCC reported in 101 primary studies. We employ meta-regression methods commonly used in economics and other fields to detect potential selective reporting in the literature. Our results suggest that, on average, the authors of primary studies tend to report preferentially estimates for which the 95% confidence interval excludes zero, which creates an upward bias in the literature. In other words, we observe that small estimates of the SCC are associated with less uncertainty (expressed as the approximate standard error used to compute the lower bound of the confidence interval) than large estimates. The finding suggests that some small estimates with large uncertainty—that is, not ruling out negative values of the SCC—might be selectively omitted from the literature. Our results also indicate that selective reporting tends to be stronger in studies published in peer-reviewed journals than in unpublished manuscripts.

Three qualifications are in order. First, we do not suggest that selective reporting in the literature on the social cost of carbon is intentional; in contrast, we believe that, as in many other fields of economics, it reflects the implicit urge to produce interesting results that are useful for policy-making: results that, in this case, help save the planet. There is an overwhelming consensus that the social costs of carbon are positive, so perhaps it makes sense to disregard estimates that are inconsistent with this view, because they probably arise from model misspecification or other estimation shortcomings. The problem is that while unintuitively small estimates are easy to recognize because of the natural lower limit of zero, there exists no obvious upper limit for the SCC. If researchers omit many small estimates but report most of the large ones (which might also be due to random misspecifications), the literature gets on average skewed toward larger estimates.

Second, we use meta-analysis methods that are designed for the synthesis of regression estimates. The estimates of the social cost of carbon are not regression-based, but mostly produced by calibrations and Monte Carlo simulations. When the authors report confidence intervals for their estimates, we argue we can use the same intuition which underlies the classical meta-analysis methods for the detection of selection reporting. Nevertheless, the large asymmetry in uncertainty about the SCC—in particular, uncertainty about potential high-impact catastrophic events triggered by climate change—leads to asymmetrical confidence intervals reported in many studies, which may, in turn, influence our estimates of the selective reporting bias. While the classical meta-analysis methods assume a symmetrical distribution of estimates, we find no evidence that the asymmetry would drive the results in our case.

Third, our results concerning selective reporting are based on a sub-sample of all available estimates of the social cost of carbon. Only about a third of the estimates are reported with a measure of uncertainty from which approximate standard errors can be computed. As an alternative, we also explore the distribution of estimates reported in studies (even if no measures of uncertainty are reported for the individual estimates), but for this exercise we can only use studies that report multiple estimates of the SCC. Both approaches produce remarkably similar results concerning the magnitude of selective reporting in the literature, but yield different estimates of the SCC corrected for the selective reporting bias: the values vary in the range 0–130 USD per ton of carbon in 2010 prices for emission year 2015. The range corresponds to the mean of median SCC values obtained by individual models or studies, not a confidence interval for the "true" SCC: especially the upper bound is difficult to pin down because of the potential catastrophic outcomes of climate change, whose probability is difficult to quantify.

References

Ackerman, F. & C. Munitz (2012): "Climate damages in the FUND model: A disaggregated analysis." *Ecological Economics* **77(C)**: pp. 219–224.

Ackerman, F. & E. A. Stanton (2012): "Climate Risks and Carbon Prices: Revising the Social Cost of Carbon." *Economics: The Open-Access, Open-Assessment E-Journal* **6(10)**: pp. 1–27.

Anthoff, D., C. Hepburn, & R. S. J. Tol (2009a): "Equity weighting and the marginal damage costs of climate change." *Ecological Economics* **68(3)**: pp. 836–849.

Anthoff, D., S. K. Rose, R. S. J. Tol, & S. Waldhoff (2011): "The Time Evolution of the Social Cost of Carbon: An Application of FUND." *Papers WP405*, Economic and Social Research Institute (ESRI).

Anthoff, D. & R. S. J. Tol (2010): "On international equity weights and national decision making on climate change." *Journal of Environmental Economics and Management* **60(1)**: pp. 14–20.

Anthoff, D. & R. S. J. Tol (2013): "The uncertainty about the social cost of carbon: A decomposition analysis using fund." *Climatic Change* **117(3)**: pp. 515–530.

Anthoff, D., R. S. J. Tol, & G. W. Yohe (2009b): "Discounting for Climate Change." *Economics: The Open-Access, Open-Assessment E-Journal* **3(24)**: pp. 1–24.

Anthoff, D., R. S. J. Tol, & G. W. Yohe (2009c): "Risk aversion, time preference, and the social cost of carbon." *Environmental Research Letters* **4**: pp. 1–7.

Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* **6(4)**: pp. 453–470.

Ayres, R. & J. Walter (1991): "The greenhouse effect: Damages, costs and abatement." *Environmental & Resource Economics* **1(3)**: pp. 237–270.

AZAR, C. (1994): "The Marginal Cost of CO2 Emissions." *Energy* **19(12)**: pp. 1255–1261.

AZAR, C. & T. STERNER (1996): "Discounting and distributional considerations in the context of global warming." *Ecological Economics* **19(2)**: pp. 169–184.

VAN DEN BIJGAART, I., R. GERLAGH, L. KORSTEN, & M. LISKI (2013): "A Simple Formula for the Social Cost of Carbon." *Working Paper Series 83*, Fondazione Eni Enrico Mattei (FEEM).

BOTH, C., A. V. ARTEMYEV, B. BLAAUW, R. J. COWIE, A. J. DEKHUIJZEN, T. EEVA, A. ENEMAR, L. GUSTAFSSON, E. V. IVANKINA, A. JARVINEN, N. B. METCALFE, N. E. I. NYHOLM, J. POTTI, P.-A. RAVUSSIN, J. J. SANZ, B. SILVERIN, F. M. SLATER, L. V. SOKOLOV, J. TOROK, , W. WINKEL, J. WRIGHT, H. ZANG, & M. E. VISSER (2004): "Large-scale geographical variation confirms that climate change causes birds to lay earlier." *Proceedings of the Royal Society of London. Series B, Biological Sciences* **271**: p. 1657–1662.

BOTZEN, W. & J. C. VAN DEN BERGH (2012): "How sensitive is Nordhaus to Weitzman? Climate policy in DICE with an alternative damage function." *Economics Letters* **117(1)**: pp. 372–374.

CAI, Y., K. L. JUDD, & T. S. LONTZEK (2012): "Open science is necessary." *Nature Climate Change* **2(5)**: p. 299.

CAI, Y., K. L. JUDD, & T. S. LONTZEK (2013): "The Social Cost of Stochastic and Irreversible Climate Change." *NBER Working Papers 18704*, National Bureau of Economic Research, Inc. (NBER).

CARD, D. & A. B. KRUEGER (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review* **85(2)**: pp. 238–43.

CERONSKY, M., D. ANTHOFF, C. HEPBURN, & R. S. J. TOL (2011): "Checking the Price Tag on Catastrophe: The Social Cost of Carbon Under Non-linear Climate Response." *Working Paper Series 392*, Economic and Social Research Institute (ESRI).

Clarkson, R. & K. Deyes (2002): "Estimating the social cost of carbon emissions." *Government Economic Service Working Papers 140*, HM Treasury, London.

Cline, W. R. (1992): *The Economics of Global Warming.* Institute for International Economics, Washington, D.C.

Cline, W. R. (1997): "Environment, Energy, and Economy." In Y. Kaya & K. Yokobori (editors), "Modelling Economically Efficient Abatement of Greenhouse Gases," chapter 3, pp. 99–122. United Nations University Press, Tokyo.

Cline, W. R. (2004): "Meeting the Challenge of Global Warming." *Copenhagen consensus challenge paper*, National Environmental Assessment Institute, Copenhagen, Denmark.

Cook, J., D. Nuccitelli, S. A. Green, M. Richardson, B. Winkler, R. Painting, R. Way, P. Jacobs, & A. Skuce (2013): "Quantifying the consensus on anthropogenic global warming in the scientific literature." *Environmental Research Letters* **8**: pp. 1–7.

Cook, J., D. Nuccitelli, A. Skuce, P. Jacobs, R. Painting, R. Honeycutt, S. A. Green, S. Lewandowsky, M. Richardson, & R. G. Wayi (2014): "Reply to 'Quantifying the consensus on anthropogenic global warming in the scientific literature: A re-analysis'." *Energy Policy* **73**: pp. 706–708.

Darling, E. S. & I. M. Côté (2008): "Quantifying the evidence for ecological synergies." *Ecology Letters* **11(12)**: pp. 1278–1286.

Dennig, F. (2013): "Inequality in Climate Change: A modification of RICE." Paper presented at 20th Annual Conference European Association of Environmental and Resource Economists (EAERE), Toulouse.

Dietz, S. (2011): "High impact, low probability? An empirical analysis of risk in the economics of climate change." *Climatic Change* **108(3)**: pp. 519–541.

Doucouliagos, H. & T. D. Stanley (2009): "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* **47(2)**: pp. 406–428.

Doucouliagos, H. & T. D. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Downing, T., D. Anthoff, R. Butterfield, M. Ceronsky, M. Grubb, J. Guo, C. Hepburn, C. Hope, A. Hunt, A. Li, A. Markandya, S. Moss, A. Nyong, R. S. J. Tol, & P. Watkiss (2005): "Social Cost of Carbon: A Closer Look at Uncertainty." *Technical report*, Department of Environment, Food and Rural Affairs (DEFRA), London.

Downing, T. E., N. Eyre, R. Greener, & D. Blackwell (1996): "Projected Costs of Climate Change for Two Reference Scenarios and Fossil Fuel Cycles." *Report to the European Commission, project ExternE*, Environmental Change Unit, Oxford.

Egger, M., G. D. Smith, M. Scheider, & C. Minder (1997): "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* **316**: pp. 629–634.

EPA & NHTSA (2009): "Proposed Rulemaking to Establish Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards." *Federal Register* **74(186)**: p. 49454–49789.

Espagne, E., B. P. Fabert, A. Pottier, F. Nadaud, & P. Dumas (2012): "Disentangling the Stern/Nordhaus Controversy: Beyond the Discounting Clash." *Working Paper Series 61*, Fondazione Eni Enrico Mattei (FEEM).

Eyre, N., T. Downing, R. Hoekstra, & K. Rennings (1999): "Externalities of Energy, Vol. 8: Global Warming." *Report to the European Commission, project ExternE*, Office for Official Publications of the European Communities, Luxembourg.

Fankhauser, S. (1994): "The Social Costs of Greenhouse Gas Emissions: An Expected Value Approach." *The Energy Journal* **15(2)**: pp. 157–184.

Foley, D. K., A. Rezai, & L. Taylor (2013): "The social cost of carbon emissions:

Seven propositions." *Economics Letters* **121(1)**: pp. 90–97.

Gerlagh, R. & M. Liski (2012): "Carbon Prices for the Next Thousand Years." *CESifo Working Paper Series 3855*, CESifo Group, Munich.

Gienapp, P., R. Leimu, & J. Merilä (2007): "Responses to climate change in avian migration time—microevolution versus phenotypic plasticity." *Climate Research* **35**: p. 25–35.

Golosov, M., J. Hassler, P. Krusell, & A. Tsyvinski (2014): "Optimal Taxes on Fossil Fuel in General Equilibrium." *Econometrica* **82(1)**: pp. 41–88.

Greenstone, M., E. Kopits, & A. Wolverton (2013): "Developing a Social Cost of Carbon for US Regulatory Analysis: A Methodology and Interpretation." *Review of Environmental Economics and Policy* **7(1)**: pp. 23–46.

Guo, J., C. J. Hepburn, R. S. Tol, & D. Anthoff (2006): "Discounting and the Social Cost of Climate Change: A Closer Look at Uncertainty." *Environmental Science & Policy* **9(3)**: pp. 205–216.

Haraden, J. (1992): "An improved shadow price for CO2." *Energy* **17(5)**: pp. 419–426.

Haraden, J. (1993): "An updated shadow price for CO2." *Energy* **18(3)**: pp. 303–307.

Havranek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85(2)**: pp. 234–244.

Havranek, T., Z. Irsova, & K. Janda (2012): "Demand for Gasoline is More Price-Inelastic than Commonly Thought." *Energy Economics* **34(1)**: p. 201–207.

Hedges, L. V. (1992): "Modeling Publication Selection Effects in Meta-Analysis." *Statistical Science* **7(2)**: pp. 246–255.

Hohmeyer, O. (1996): "Social Costs of Climate Change: Strong Sustainability and Social Costs." In O. Hohmeyer, R. Ottinger, & K. Rennings (editors), "Social Costs and Sustainability: Valuation and Implementation in the Energy and

Transport Sector," pp. 61–83. Springer, Berlin.

Hohmeyer, O. (2004): "Verguetung nach dem EEG: Subvention oder fairer Ausgleich externer Kosten?" In H. Ziesing (editor), "Externe Kosten in der Stromerzeugung," pp. 11–24. Frankfurt am Main: VWEW Energieverlag.

Hohmeyer, O. & M. Gaertner (1992): *The Costs of Climate Change - A Rough Estimate of Orders of Magnitude.* Fraunhofer-Institut fur Systemtechnik und Innovationsforschung, Karlsruhe.

Hope, C. W. (2005a): "Exchange Rates and the Social Cost of Carbon." *Working Paper Series 5*, Judge Institute of Management, Cambridge, UK.

Hope, C. W. (2005b): "The Climate Change Benefits of Reducing Methane Emissions." *Climatic Change* **68(1-2)**: pp. 21–39.

Hope, C. W. (2006): "The Marginal Impact of CO2 from PAGE2002: An Integrated Assessment Model Incorporating the IPCC's Five Reasons for Concern." *Integrated Assessment Journal* **6(1)**: pp. 19–56.

Hope, C. W. (2008a): "Discount rates, equity weights and the social cost of carbon." *Energy Economics* **30(3)**: pp. 1011–1019.

Hope, C. W. (2008b): "Optimal Carbon Emissions and the Social Cost of Carbon over Time under Uncertainty." *Integrated Assessment Journal* **8(1)**: pp. 107–122.

Hope, C. W. (2011): "The social cost of CO2 from the PAGE09 model." *Economics Discussion Papers 39*, Kiel Institute for the World Economy.

Hope, C. W. & P. Maul (1996): "Valuing the impact of CO2 emissions." *Energy Policy* **24(3)**: pp. 211–219.

Howarth, R. B., M. D. Gerst, & M. E. Borsuk (2014): "Risk mitigation and the social cost of carbon." *Global Environmental Change* **24**: pp. 123–131.

Hwang, I., F. Reynes, & R. Tol (2013): "Climate Policy Under Fat-Tailed Risk: An Application of DICE." *Environmental & Resource Economics* **56(3)**: pp. 415–436.

IPCC (1995): *Intergovernmental Panel on Climate Change Second Assessment Re-*

port: Climate Change 1995. Working Group II: Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses. Cambridge University Press, UK.

IPCC (2007): *Intergovernmental Panel on Climate Change Fourth Assessment Report: Climate Change 2007. Working Group II: Impacts, Adaptations and Vulnerability.* Cambridge University Press, UK and NY.

IPCC (2014): *Intergovernmental Panel on Climate Change Fifth Assessment Report: Climate Change 2014. Working Group II: Impacts, Adaptations and Vulnerability.* Cambridge University Press.

IWG (2010): "Technical Support Document: Social Cost of Carbon for Regulatory Impact Analysis." *Technical report*, U.S. Government.

IWG (2013): "Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis." *Technical report*, U.S. Government.

Jensen, S. & C. P. Traeger (2014a): "Optimal climate change mitigation under long-term growth uncertainty: Stochastic integrated assessment and analytic findings." *European Economic Review* **69(C)**: pp. 104–125.

Jensen, S. & C. P. Traeger (2014b): "Optimally Climate Sensitive Policy under Uncertainty and Learning." Paper presented at 2014 Annual Conference of the American Economic Association (AEA), Philadelphia.

Johnson, L. T. & C. Hope (2012): "The social cost of carbon in U.S. regulatory impact analyses: An introduction and critique." *Journal of Environmental Studies and Sciences* **2(3)**: pp. 205–221.

Kemfert, C. & W.-P. Schill (2010): "Methane Mitigation." In B. Lomborg (editor), "Smart Solutions to Climate Change," pp. 172–197. Cambridge University Press, Cambridge.

Kopp, R. E., A. Golub, N. O. Keohane, & C. Onda (2012): "The Influence of the Specification of Climate Change Damages on the Social Cost of Carbon." *Economics: The Open-Access, Open-Assessment E-Journal* **6(13)**: pp. 1–40.

Krakovsky, M. (2004): "Register of Perish." *Scientific American* **291**: pp. 18–20.

Lemoine, D. & C. Traeger (2014): "Watch Your Step: Optimal Policy in a Tipping Climate." *American Economic Journal: Economic Policy* **6(1)**: pp. 137–66.

Link, P. M. & R. S. J. Tol (2004): "Possible Economic Impacts of a Shutdown of the Thermohaline Circulation: An Application of FUND." *Portuguese Economic Journal* **3**: pp. 99–114.

Lintunen, J. & L. Vilmi (2013): "On optimal emission control: Taxes, substitution and business cycles." *Research Discussion Papers 24*, Bank of Finland.

Maclean, I. M. D. & R. J. Wilson (2011): "Recent ecological responses to climate change support predictions of high extinction risk." *Proceedings of the National Academy of Sciences of the United States of America* **108(30)**: p. 12337–12342.

Maddison, D. (1995): "A cost-benefit analysis of slowing climate change." *Energy Policy* **23(4-5)**: pp. 337–346.

Manne, A. (2004): "Climate Change: An Opponent's Notes." In B. Lomborg (editor), "Global Crises, Global Solutions," pp. 49–55. Cambridge University Press, New York.

Marten, A. L. & S. C. Newbold (2012): "Estimating the social cost of non-CO2 GHG emissions: Methane and nitrous oxide." *Energy Policy* **51(C)**: pp. 957–972.

Massad, T. J. & L. A. Dyer (2010): "A meta-analysis of the effects of global environmental change on plant-herbivore interactions." *Arthropod-Plant Interactions* **4(3)**: p. 181–188.

Mendelsohn, R. (2004): "Climate Change: An Opponent's Notes." In B. Lomborg (editor), "Global Crises, Global Solutions," pp. 44–48. Cambridge University Press, New York.

Menzel, A., T. H. Sparks, N. Estrella, E. Koch, A. Aasa, R. Ahas, K. Alm-Kubler, P. Bissolli, O. Braslavská, A. Briede, F. M. Chmielewski, Z. Crepinsek, Y. Curnel, A. Dahl, C. Defila, A. Donnelly, Y. Filella, K. Jatczak, F. Mage, A. Mestre, O. Nordli, J. Penuelas, P. Pirinen,

V. Remišová, H. Scheifinger, M. Striz, A. Susnik, A. J. H. van Vliet, F.-E. Wielgolaski, S. Zach, & A. Zust (2006): "European phenological response to climate change matches the warming pattern." *Global Change Biology* **12(10)**: p. 1969–1976.

Michaels, J. P. (2008): "Evidence for "Publication Bias" Concerning Global Warming in Science and Nature." *Energy & Environment* **19(2)**: pp. 287–301.

Moyer, E. J., M. D. Woolley, M. Glotter, & D. A. Weisbach (2013): "Climate Impacts on Economic Growth as Drivers of Uncertainty in the Social Cost of Carbon." *Working Paper Series 02*, Center for Robust Decision Making on Climate & Energy Policy (RDCEP), University of Chicago.

Narita, D., D. Anthoff, & R. S. J. Tol (2009): "Damage Costs of Climate Change through Intensification of Tropical Cyclone Activities: An Application of FUND." *Climate Research* **39**: pp. 87–97.

Narita, D., R. S. J. Tol, & D. Anthoff (2010): "Economic costs of extratropical storms under climate change: an application of FUND." *Journal of Environmental Planning and Management* **53(3)**: pp. 371–384.

Necker, S. (2014): "Scientific misbehavior in economics." *Research Policy* **(forthcoming)**.

Newbold, S. C., C. Griffiths, C. Moore, A. Wolverton, & E. Kopits (2013): "A Rapid Assessment Model for Understanding the Social Cost of Carbon." *Climate Change Economics* **04(01)**: p. 1350001.

Newbold, S. C. & A. L. Marten (2014): "The value of information for integrated assessment models of climate change." *Journal of Environmental Economics and Management* **68(1)**: pp. 111–123.

Newell, R. G. & W. A. Pizer (2003): "Discounting the distant future: how much do uncertain rates increase valuations?" *Journal of Environmental Economics and Management* **46(1)**: pp. 52–71.

Nordhaus, W. (1982): "How Fast Should We Graze the Global Commons?" *Amer-*

ican Economic Review **72(2)**: pp. 242–46.

Nordhaus, W. & J. Boyer (2000): *Warming the World: Economic Models of Global Warming.* The MIT Press, Cambridge.

Nordhaus, W. & P. Sztorc (2014): "DICE 2013: Introduction and users manual." Users manual.

Nordhaus, W. D. (1991): "To Slow or Not to Slow: The Economics of the Greenhouse Effect." *Economic Journal* **101(407)**: pp. 920–37.

Nordhaus, W. D. (1993): "Rolling the 'DICE': an optimal transition path for controlling greenhouse gases." *Resource and Energy Economics* **15(1)**: pp. 27–50.

Nordhaus, W. D. (1994): *Managing the Global Commons: The Economics of Climate Change.* The MIT Press, Cambridge.

Nordhaus, W. D. (2008): *A Question of Balance: Weighing the Options on Global Warming Policies.* Yale University Press, New Haven.

Nordhaus, W. D. (2010): "Economic aspects of global warming in a post-Copenhagen environment." In "Proceedings of the National Academy of Sciences of the United States of America," volume 107, pp. 11721–11726.

Nordhaus, W. D. (2011): "Estimates of the Social Cost of Carbon: Background and Results from the RICE-2011 Model." *NBER Working Papers 17540*, National Bureau of Economic Research, Inc. (NBER).

Nordhaus, W. D. & D. Popp (1997): "What is the Value of Scientific Knowledge? An Application to Global Warming Using the PRICE Model." *The Energy Journal* **18(1)**: pp. 1–45.

Nordhaus, W. D. & Z. Yang (1996): "A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies." *American Economic Review* **86(4)**: pp. 741–65.

NRC (2009): *Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use.* National Research Council of the National Academies, National Academies Press.

PARMESAN, C. (2007): "Influences of species, latitudes and methodologies on estimates of phenological response to global warming." *Global Change Biology* **13**: p. 1860–1872.

PARRY, I. W. H. (1993): "Some estimates of the insurance value against climate change from reducing greenhouse gas emissions." *Resource and Energy Economics* **15(1)**: pp. 99–115.

PEARCE, D. (2003): "The Social Cost of Carbon and its Policy Implications." *Oxford Review of Economic Policy* **19(3)**: pp. 362–384.

PECK, S. C. & T. J. TEISBERG (1993): "Global warming uncertainties and the value of information: An analysis using CETA." *Resource and Energy Economics* **15(1)**: pp. 71–97.

PENNER, S., J. HARADEN, & S. MATES (1992): "Long-term global energy supplies with acceptable environmental impacts." *Energy* **17(10)**: pp. 883–899.

PERRISSIN-FABERT, B., P. DUMAS, & J.-C. HOURCADE (2012): "What Social Cost of Carbon? A Mapping of the Climate Debate." *Working Paper Series 34*, Fondazione Eni Enrico Mattei (FEEM).

PINDYCK, R. S. (2013): "Climate Change Policy: What Do the Models Tell Us?" *Journal of Economic Literature* **51(3)**: pp. 860–72.

PLAMBECK, E. L. & C. HOPE (1996): "PAGE95 : An updated valuation of the impacts of global warming." *Energy Policy* **24(9)**: pp. 783–793.

PYCROFT, J., L. VERGANO, & C. HOPE (2014): "The economic impact of extreme sea-level rise: Ice sheet vulnerability and the social cost of carbon dioxide." *Global Environmental Change* **24**: p. 99–107.

PYCROFT, J., L. VERGANO, C. W. HOPE, D. PACI, & J. C. CISCAR (2011): "A Tale of Tails: Uncertainty and the Social Cost of Carbon Dioxide." *Economics: The Open-Access, Open-Assessment E-Journal* **5(22)**: pp. 1–29.

REILLY, J. & K. RICHARDS (1993): "Climate change damage and the trace gas index issue." *Environmental & Resource Economics* **3(1)**: pp. 41–61.

Rezai, A. & F. van der Ploeg (2014): "Abandoning Fossil Fuel; How fast and how much?" *OxCarre Working Papers 123*, Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford.

Rosenthal, R. (1979): "The 'File Drawer Problem' and Tolerance for Null Results." *Psychological Bulletin* **86**: pp. 638–41.

Roughgarden, T. & S. H. Schneider (1999): "Climate change policy: quantifying uncertainties for damages and optimal carbon taxes." *Energy Policy* **27(7)**: pp. 415–429.

Rusnak, M., T. Havranek, & R. Horvath (2013): "How to Solve the Price Puzzle? A Meta-Analysis." *Journal of Money, Credit and Banking* **45(1)**: pp. 37–70.

Schauer, M. (1995): "Estimation of the greenhouse gas externality with uncertainty." *Environmental & Resource Economics* **5(1)**: pp. 71–82.

Siegfried, J. J. (2012): "Minutes of the Meeting of the Executive Committee: Chicago, IL, January 5, 2012." *American Economic Review* **102(3)**: pp. 645–52.

Sohngen, B. (2010): "Forestry Carbon Sequestration." In B. Lomborg (editor), "Smart Solutions to Climate Change," pp. 114–132. Cambridge University Press, Cambridge.

Stanley, T., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh, & R. S. R. and (2013): "Meta-Analysis Of Economics Research Reporting Guidelines." *Journal of Economic Surveys* **27(2)**: pp. 390–394.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stern, N., S. Peters, V. Bakhshi, A. Bowen, C. Cameron, S. Catovsky, D. Crane, S. Cruickshank, S. Dietz, N. Edmonson, S.-L. Garbett,

L. Hamid, G. Hoffman, D. Ingram, B. Jones, N. Patmore, H. Radcliffe, R. Sathiyarajah, M. Stock, C. Taylor, T. Vernon, H. Wanjie, & D. Zenghelis (2006): *Stern Review: The Economics of Climate Change*. Cambridge University Press, New York.

Stern, N. & C. Taylor (2007): "Climate Change: Risk, Ethics, and the Stern Review." *Nature* **317**: pp. 203–204.

Swanson, K. L. (2013): "Emerging selection bias in large-scale climate change simulations." *Geophysical Research Letters* **40(12)**: p. 3184–3188.

Tol, R. S. J. (1995): "The Damage Costs of Climate Change Toward More Comprehensive Calculations." *Environmental and Resource Economics* **5**: pp. 353–374.

Tol, R. S. J. (1999): "The Marginal Costs of Greenhouse Gas Emissions." *The Energy Journal* **20(1)**: pp. 61–81.

Tol, R. S. J. (2002a): "Estimates of the Damage Costs of Climate Change. Part I: Benchmark Estimates." *Environmental & Resource Economics* **21(1)**: pp. 47–73.

Tol, R. S. J. (2002b): "Estimates of the Damage Costs of Climate Change. Part II: Dynamic Estimates." *Environmental & Resource Economics* **21(2)**: pp. 135–160.

Tol, R. S. J. (2005a): "Emission abatement versus development as strategies to reduce vulnerability to climate change: an application of FUND." *Environment and Development Economics* **10(05)**: pp. 615–629.

Tol, R. S. J. (2005b): "The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties." *Energy Policy* **33(16)**: p. 2064–2074.

Tol, R. S. J. (2008): "The Social Cost of Carbon: Trends, Outliers and Catastrophes." *Economics - The Open-Access, Open-Assessment E-Journal* **2(25)**: pp. 1–22.

Tol, R. S. J. (2010): "Carbon Dioxide Mitigation." In B. Lomborg (editor), "Smart Solutions to Climate Change," pp. 74–105. Cambridge University Press, Cambridge.

Tol, R. S. J. (2011): "The Social Cost of Carbon." *Annual Review of Resource*

Economics **3(1)**: pp. 419–443.

Tol, R. S. J. (2012): "On the Uncertainty About the Total Economic Impact of Climate Change." *Environmental & Resource Economics* **53(1)**: pp. 97–116.

Tol, R. S. J. (2013a): "Climate policy with Bentham–Rawls preferences." *Economics Letters* **118(3)**: pp. 424–428.

Tol, R. S. J. (2013b): "Targets for global climate policy: An overview." *Journal of Economic Dynamics and Control* **37(5)**: pp. 911–928.

Tol, R. S. J. (2014): "Quantifying the consensus on anthropogenic global warming in the literature: A re-analysis." *Energy Policy* **73**: pp. 701–705.

Tol, R. S. J. & T. Downing (2001): "The marginal costs of climate changing emissions." In Friedrich & Bickel (editors), "Environmental External Costs of Transport," Springer Verlag Heidelberg.

Uzawa, H. (2003): *Economic Theory and Global Warming.* Cambridge University Press, Cambridge.

Wahba, M. & C. Hope (2006): "The marginal impact of carbon dioxide under two scenarios of future emissions." *Energy Policy* **34(17)**: pp. 3305–3316.

Waldhoff, S., D. Anthoff, S. Rose, & R. S. J. Tol (2011): "The Marginal Damage Costs of Different Greenhouse Gases: An Application of FUND." *Economics Discussion Paper 43*, Kiel Institute for the World Economy.

Weitzman, M. (2013): "Tail-Hedge Discounting and the Social Cost of Carbon." *Journal of Economic Literature* **51(3)**: pp. 873–82.

7.A Included Studies and Summary Statistics

Table 7.8: List of studies used in the meta-analysis

Ackerman & Munitz (2012)	Haraden (1993)	Nordhaus (1994)
Ackerman & Stanton (2012)	Hohmeyer & Gaertner (1992)	Nordhaus & Yang (1996)
Anthoff *et al.* (2009a)	Hohmeyer (1996)	Nordhaus & Popp (1997)
Anthoff *et al.* (2009b)	Hohmeyer (2004)	Nordhaus & Boyer (2000)
Anthoff *et al.* (2009c)	Hope & Maul (1996)	Nordhaus (2008)
Anthoff & Tol (2010)	Hope (2005a)	Nordhaus (2010)
Anthoff *et al.* (2011)	Hope (2005b)	Nordhaus (2011)
Anthoff & Tol (2013)	Hope (2006)	Nordhaus & Sztorc (2014)
Ayres & Walter (1991)	Hope (2008a)	Parry (1993)
Azar (1994)	Hope (2008b)	Pearce (2003)
Azar & Sterner (1996)	Hope (2011)	Peck & Teisberg (1993)
van den Bijgaart *et al.* (2013)	Howarth *et al.* (2014)	Penner *et al.* (1992)
Botzen & van den Bergh (2012)	Hwang *et al.* (2013)	Perrissin-Fabert *et al.* (2012)
Cai *et al.* (2012)	Jensen & Traeger (2014a)	Plambeck & Hope (1996)
Cai *et al.* (2013)	Jensen & Traeger (2014b)	Pycroft *et al.* (2011)
Ceronsky *et al.* (2011)	Johnson & Hope (2012)	Pycroft *et al.* (2014)
Clarkson & Deyes (2002)	Kemfert & Schill (2010)	Reilly & Richards (1993)
Cline (1992)	Kopp *et al.* (2012)	Rezai & van der Ploeg (2014)
Cline (1997)	Lemoine & Traeger (2014)	Roughgarden & Schneider (1999)
Cline (2004)	Link & Tol (2004)	Schauer (1995)
Dennig (2013)	Lintunen & Vilmi (2013)	Sohngen (2010)
Dietz (2011)	Maddison (1995)	Stern *et al.* (2006)
Downing *et al.* (1996)	Manne (2004)	Stern & Taylor (2007)
Downing *et al.* (2005)	Marten & Newbold (2012)	Tol (1999)
EPA & NHTSA (2009)	Mendelsohn (2004)	Tol & Downing (2001)
Espagne *et al.* (2012)	Moyer *et al.* (2013)	Tol (2005a)
Eyre *et al.* (1999)	Narita *et al.* (2009)	Tol (2010)
Fankhauser (1994)	Narita *et al.* (2010)	Tol (2012)
Foley *et al.* (2013)	Newbold *et al.* (2013)	Tol (2013a)
Gerlagh & Liski (2012)	Newbold & Marten (2014)	Uzawa (2003)
Golosov *et al.* (2014)	Newell & Pizer (2003)	Wahba & Hope (2006)
Greenstone *et al.* (2013)	Nordhaus (1982)	Waldhoff *et al.* (2011)
Guo *et al.* (2006)	Nordhaus (1991)	Weitzman (2013)
Haraden (1992)	Nordhaus (1993)	

Notes: The last study was added on August 1, 2014.

Table 7.9: Summary statistics, estimates with standard errors

Variable	Description	Obs.	Mean	Std. dev.
SCC	The reported estimate of the social cost of carbon in USD per ton of carbon (normalized to 2015 emission year in 2010 dollars).	267	411	521
Standard error	The approximate standard error of the estimate computed from the reported lower bound of the confidence interval.	267	162	235

Continued on next page

Table 7.9: Summary statistics, estimates with standard errors (continued)

Variable	Description	Obs.	Mean	Std. dev.
Upper SE	The approximate standard error of the estimate computed from the reported upper bound.	267	1182	1921
Reviewed	= 1 if the study was published in a peer-reviewed outlet.	267	0.94	0.24
Publication year	The year of publication of the study (base: 1982).	267	27.9	4.88
Mean estimate	= 1 if the reported SCC estimate is the mean of the distribution.	267	0.30	0.46
Median estimate	= 1 if the reported SCC estimate is the median of the distribution.	267	0.64	0.48
Marginal costs	= 1 if the study estimates marginal damage costs (damage from an additional ton of carbon emitted) rather than average costs (the total impact divided by the total emissions of carbon).	267	1.00	0.06
Dynamic impacts	= 1 if the study examines dynamic impacts of climate change or uses a dynamic model of vulnerability.	267	0.12	0.32
Scenarios	= 1 if the study uses climate and economic scenarios that are internally consistent. A few studies use arbitrary assumptions about climate change.	267	0.96	0.19
FUND	= 1 if the authors use the FUND model or derive their model from FUND.	267	0.13	0.34
DICE or RICE	= 1 if the authors use the DICE/RICE model or derive their model from DICE/RICE.	267	0.69	0.46
PAGE	= 1 if the authors use the PAGE model or derive their model from PAGE.	267	0.32	0.47
PRTP	The pure rate of time preference assumed in the estimation.	217	1.12	1.54
Equity weights	= 1 if equity weighting is applied.	267	0.15	0.36
Pigovian tax	= 1 if the estimate is computed along a trajectory of emissions in which the marginal costs of emission reduction equal the SCC, then the estimate corresponds to a Pigovian tax.	267	0.57	0.50
Citations	= The logarithm of the number of Google Scholar citations of the study.	267	3.25	0.92
Journal rank	= SciMago journal rank based on the impact factor extracted from Scopus.	267	0.48	0.86

Notes: Data are collected from studies estimating the social cost of carbon. The data set is available at `meta-analysis.cz/scc`.

Table 7.10: Summary statistics, study-level medians

Variable	Description	Obs.	Mean	Std. dev.
SCC	The reported estimate of the social cost of carbon in USD per ton of carbon (normalized to 2015 emission year in 2010 dollars).	68	201	344
Standard error	The approximate standard error of the estimate computed from the reported lower bound of the confidence interval.	68	93	184
Upper SE	The approximate standard error of the estimate computed from the reported upper bound.	68	237	327
Reviewed	= 1 if the study was published in a peer-reviewed outlet.	68	0.72	0.45
Publication year	The year of publication of the study (base: 1982).	68	24.5	7.49
Mean estimate	= 1 if the reported SCC estimate is the mean of the distribution.	68	0.30	0.46
Median estimate	= 1 if the reported SCC estimate is the median of the distribution.	68	0.09	0.29
Marginal costs	= 1 if the study estimates marginal damage costs (damage from an additional ton of carbon emitted) rather than average costs (the total impact divided by the total emissions of carbon).	68	0.91	0.29
Dynamic impacts	= 1 if the study examines dynamic impacts of climate change or uses a dynamic model of vulnerability.	68	0.37	0.49
Scenarios	= 1 if the study uses climate and economic scenarios that are internally consistent. A few studies use arbitrary assumptions about climate change.	68	0.76	0.43
FUND	= 1 if the authors use the FUND model or derive their model from FUND.	68	0.31	0.47
DICE or RICE	= 1 if the authors use the DICE/RICE model or derive their model from DICE/RICE.	68	0.29	0.46
PAGE	= 1 if the authors use the PAGE model or derive their model from PAGE.	68	0.24	0.43
PRTP	The pure rate of time preference assumed in the estimation.	53	1.44	1.01
Equity weights	= 1 if equity weighting is applied.	68	0.19	0.39
Pigovian tax	= 1 if the estimate is computed along a trajectory of emissions in which the marginal costs of emission reduction equal the SCC, then the estimate corresponds to a Pigovian tax.	68	0.20	0.40
Citations	= The logarithm of the number of Google Scholar citations of the study.	68	3.35	1.49
Journal rank	= SciMago journal rank based on the impact factor extracted from Scopus.	68	1.62	2.74

Notes: Data are collected from studies estimating the social cost of carbon. The data set is available at `meta-analysis.cz/scc`.

Appendix A

Response to Reviewers

Pre-defense report on manuscript "Six Essays on Meta-Regression Analysis" as of December 3rd, 2014.

I thank the reviewers for insightful comments on the pre-defense version of my dissertation. Since the reviewers suggest that the dissertation can be submitted without major changes, I only make minor adjustments in the text.

Response to Comments from Jan Babecky Dr. Babecky has one suggestion for revision: following his recommendation I include a section summarizing the lessons learned for conducting meta-regression analysis in economics. I include the section at the end of Introduction and, following Dr. Babecky's suggestion, structure the discussion according to the following topics: selection of primary studies, testing for publication bias, selection of explanatory variables, robustness checks, and best-practice estimation.

Response to Comments from Tomas Cahlik I am grateful to Prof. Cahlik for his kind words on my dissertation. He does not suggest any revisions.

Response to Comments from Oldrich Dedek I thank my advisor, Prof. Dedek, for his kind assessment of my work. He does not suggest any revisions.

Response to Comments from Jarko Fidrmuc I am grateful to Prof. Fidrmuc for his kind words on my dissertation. He does not suggest any revisions.