Charles University

Faculty of Social Sciences
Institute of Economic Studies

DISSERTATION

Essays in Applied Meta-Analysis

Author: **Mgr. Petr Polák, MSc.**

Supervisor: **doc. PhDr. Tomáš Havránek, Ph.D.**

Academic Year: **2020/2021**

Abstract

The dissertation consists of three papers presenting applications of meta-analysis in economics and introductory chapter which discusses the development of meta-analysis in author's perspective as well as strengths and weaknesses of this quantitative method to synthesize empirical research.

In the first paper a meta-analysis of the impact of information and communication technology (ICT) on economic is carried out. The ICT performance has been the subject of academic research for several decades, and despite the remarkable and significant innovation in computer technology, usage, and investments, only a small growth in productivity has been observed. This observations has been coined the productivity paradox. This paper uses meta-analytical methods to examine publication bias and the size of the output elasticity of ICT capital. The empirical part is based on a collection of more than 800 estimates of ICT payoff effects from more than 70 studies written in the last 20 years. The meta-analysis reveals a strong presence of publication bias within the ICT productivity literature and, using a mixed effect multilevel model, estimates the output ICT elasticity to be only 0.3%, which is more than ten times smaller than what was reported by a previous meta-analysis from 10 years ago.

The second paper focuses on the trade effect of the euro, which was estimated by many studies, but their results vary greatly. This meta-analysis collects 3,323 estimates of the euro effect along with 28 characteristics of estimation design from almost 60 studies and quantitatively examines the literature. The results show evidence of publication bias, but they also suggest that the bias decreases over time. After correcting for the bias, the meta-analysis shows that the literature is consistent with an effect ranging between 2 and 6 %. The results from Bayesian model averaging, which takes into account model uncertainty, show that the differences among estimates are systematically driven by data sources, data structure, control variables, and estimation techniques. The mean reported estimate of the euro's trade effect conditional on best-practice approach is 3 %, but is not statistically different from zero.

In the third paper we use more than 1600 estimates from 71 studies to investigate the relation between international trade flows and trade costs, which are mainly tariffs. Our results suggest that the empirical literature suffers from the presence of publication bias, which has exaggerated the effect (the elasticity is higher). After accounting for publication bias, we estimate trade elasticity

with respect to trade cost to be between -0.9 and -2.0. We also identified several properties that explain the heterogeneity of the reported estimates. The results from Bayesian model averaging, which takes into account model uncertainty, show that the differences among estimates are systematically driven by the data source, control variables, and estimation techniques. The mean reported estimate of the trade cost conditional on the best-practice approach is not statistically different from zero due to the large confidence interval.

Thank you

It took me a long time, but I am grateful that I have the possibility to present this thesis. This would not be possible without the support and approach of my advisor - Tomáš Havránek. I was lucky that I met him during my undergraduate studies, and since then, he has been my guarding angel. His enthusiasm, ability to formulate ideas, and work morale was a source of inspiration and motivation. Thank you.

Further, I would like to thank my colleagues at the Institute of Economic Studies as well as colleagues at the Czech National Bank for their comments and inspiration. Many of them are and were my fellow PhD students, and the environment helped me a lot.

I would also like to express my gratitude to the Institute of Economic Studies for providing me the opportunity to follow an academic career and pass the knowledge by allowing me to teach over the years quite extensively.

Finally, I would like to dedicate this dissertation to my wife and children. They allowed me to pursue this goal instead of spending my time with them.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Polák, Petr: *Essays in Applied Meta-Analysis.* DISSERTATION . Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 178. Advisor: doc. PhDr. Tomáš Havránek, Ph.D.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> The replication of research is an essential component of scientific methodology. Only through replication of the results of others can scientists unify the disparate findings of various researchers in a discipline into a defensible, consistent, coherent body of knowledge (Dewald *et al.*, 1986, p. 600).

This dissertation contains three separate papers with a focus on different topics in macroeconomics. Nonetheless, all of them build upon the same underlying phenomenon – meta-analysis and publication bias. They all use very similar methodologies, but one can also see the development of the techniques used for the analysis. The first two papers (and chapters) are solo-authored, the last one I wrote jointly with my colleagues. All three papers analyze quite well-discussed topics in the economics: *Productivity paradox*, *Effect of common currency on bilateral trade* and *Effect of trade costs on bilateral trade*. The first paper was published in *Information Economics and Policy*, while the second one was published in *Journal of Economic Surveys*. The third paper in the form of a draft has been submitted to the working paper series so far.

1.1 Meta-analysis is. . .

A meta-analysis is a tool or approach to synthesize empirical findings of existing literature systematically. In his pioneering work Glass (1976, p.1) defined "meta-analysis" as the statistical "analysis of analyses", a tool for integrating findings from the collection of individual studies. For academia, it is essential to summarize results from previous research focused on one topic, collect and connect all the pieces of evidence that are spread in the available literature, understandably present them, and explore factors that influence research results. This should be done in the most objective and unbiased way possible, and precisely that is the basic idea behind meta-analysis. Nevertheless, following Glass (1976) we might end up with a systematic review, which is basically a summary of previous studies. The difference between meta-analyses and systematic reviews is in the quantitative approach. We can follow the broad definition from Normand (1999, p. 321) that meta-analysis is "the quantitative review and synthesis of the results of related but independent studies" and then continue with Petitti (2001, p. 3625), who states that "not all systematic reviews include a quantitative synthesis". In general, we can formulate meta-analysis targets as evaluating and reviewing the results of empirical research and providing a systematic and comprehensive summary.

The importance of meta-analysis is growing with the increasing number of published empirical studies. Narrative reviews are no longer able to capture and present this enormous amount of data in a comprehensive way. Meta-analysis offers a set of econometric tools specially designed for this purpose. Stanley (2001, p. 131) sees meta-analysis as "a body of statistical methods that have been found useful in reviewing and evaluating empirical research results." The basic idea is to examine factors that influence research results of some phenomena. The dependent variable is the effect size estimate, while the independent variables consist of various information about each study like data characteristics, the method used for analysis, sample size, sometimes even the researcher's occupation or employer.

Studies provide estimates of different sizes, and in cases when the variance of the results is too large to be justified by the disturbance terms, we speak about between-study heterogeneity. We try to explain this heterogeneity by specific differences between the studies. Thus, we code properties of the studies into variables, and later on, we test for the presence of heterogeneity. Christensen (2003) describes that there are two general types of heterogeneity present in the research: factual and methodological. Factual heterogeneity concerns real differences in the effect due to actual differences in the tested sample, such as when a study about the same phenomena was conducted at a different time or in a different country. We can take the first topic of information and communication technology as an example; there could be a difference between developing countries, where the economy is based on manufacturing, and developed countries with a service-oriented economy. Methodological heterogeneity is the result of different study approaches; it could be models used, data characteristics, econometric methods, or control variables used. The meta-analysis attempts to take all the above-discussed factors (data, estimation, and publication) into account and evaluate if they play a significant role in the reported empirical estimates' heterogeneity.

When we want to determine which study characteristics influence the results, we might overlook an important factor that might cause a significant difference in the outcome – the researcher and other people, which are part of the publication process. We call this a publication bias. Stanley (2005) has for the publication bias two categories. If the researcher's primary motivation is to get published, findings contradicting previous studies or such that conflict with the theory may get concealed. In many cases, different model specifications produce an utterly different outcome, thus modifying the model or data until acceptable results can occur. Any modification (e.g., restraining dataset or model modification) to obtain results aligned with the theory is labeled as publication bias of the first type. Adjusting models may also happen if a standard procedure has an insignificant outcome, and it is not the deserved outcome. Therefore, the second type of publication bias can be described as reaching the statistical significance no matter what the effect size would be. Most of the time, these two are combined. It should be stressed out that most researchers' decisions are entirely rational and justified at micro-level (e.g. changing specification when the estimated coefficient is clearly at odds with theory or elimination of possible outliers from the dataset). The downside is that these decisions cause the bias of the aggregate mean of the published estimates available in the literature.

Tampering of results needs to be measured and taken into account when conducting empirical summaries. Saying that such practices are common in all areas of research would be too strong, but we can formulate a similar hypothesis, which can be tested and accepted or rejected. The hypothesis that a large part of economic studies is affected by publication or selection bias is tested and supported by Doucouliagos & Stanley (2008) in their work based on 65 distinct pieces of empirical economics literature, involving approximately two thousand separate empirical studies. Their work confirmed that publication bias had been a serious issue in the empirical economics research (Card & Krueger, 1995; Ashenfelter & Greenstone, 2004; Stanley, 2005). Meta-analysis is a great tool to investigate the publication bias as it collects all available empirical evidence. Ioannidis *et al.* (2017) survey 159 meta-analyses made in economics and investigate the statistical power and bias of more than 64,000 estimates collected by those works. They argue that nearly 80% of reported effect sizes are exaggerated which and the reason is that authors, editors, and referees prefer statistically significant results with an intuitive sign because the economists respond to the main incentive they have – to publish or perish. According to Necker (2014), economists in her survey fell pressure to publish, and one third admit to search for control variables to get desired results or present the empirical findings selectively so that they confirm the story. Christensen & Miguel (2018) provides evidence about seeking statistical significance and discusses new practices that should promote the transparency and replicability of empirical research. At least, econometric methods can discover the presence of the publication bias and estimate the true effect beyond (Hunter & Schmidt, 2004). The approach, however, has some limits as pointed by Alinaghi & Reed (2018) and followed by Stanley & Doucouliagos (2019), who doubts that it is fully possible.

The meta-analytic approach aims to be as objective as possible, but subjectivity can never be eliminated, especially when the researcher determines which and how the dependent variables will be collected and which models for analysis will be used. It should be noted that a meta-analysis requires a sufficient number of primary studies that focus on the same topic. That is one of the weak points of meta-analysis. It would be, therefore, beneficial if there would be more replication studies. Furthermore, as Burman *et al.* (2010) points out, empirical economic research is often prone to error, and replication studies are valuable. Unfortunately, replication encounters resistance in economics (Duvendack *et al.*, 2017). Meta-analysis and replication studies can largely benefit from each other. As Duvendack *et al.* (2015) suggests, "replication studies can take the results of meta-analyses and investigate whether changing the empirical design of a study has the effect predicted by meta-analysis. conversely, replication studies may identify study characteristics that meta-analyses can incorporate in subsequent meta-regression research." I would add that replication studies make the meta-regression research more robust because it increases the sample of available estimates.

From the meta-analytic perspective, all three papers in this thesis identify the presence of significant publication bias in the examined part of economic literature. According to Ioannidis *et al.* (2017), one of the possible remedies is to promote replication and share code and data. I would agree that more transparency is better for the whole academic community. For the revision of the second paper *The Euro Effect: A Meta-analysis*, I try to collect explanatory variable that would capture if authors' publish they data or code and only 1 study out of 57 did so. On the other hand, even code and data sharing would not ultimately produce unbiased results since the author arbitrarily chooses many steps during the whole process. However, it is necessary to support alternative replications approaches that would provide similar or completely different results. I am glad that my supervisor Tomáš Havránek promotes meta-analysis transparency by publishing the code and data for all the works on meta-analysis.cz.

Another limitation of meta-analysis is that the meta-analysts have limited access to the primary studies' estimation process – they can evaluate only what is written down. The meta-analyst code more and more factors of primary studies, typically a few dozens nowadays. Not each deviation of the study design can be coded because it would be impossible to make a regression. It also might happen that the study design of the primary research has some flaws, and meta-analysts might not identify them. From the practical point, especially in macroeconomics, many studies used the same data source, and hence the results of these studies are correlated and it is very hard to tackle such issue. We should not forget about the current issue that meta-analysts are facing – having many potential explanatory variables at hand leads to an important question: how the final model will be created so that only relevant explanatory variables are included and avoid a publication selection process again.

The empirical review of academic literature is not only beneficial for academics, but it is crucial in the policy application. Politics make decisions based on the results of different analyses. If the outcomes from different sources are contradicting, the final decision might not be optimal. The results itself do not need to be contradicting per se; they just might be insignificant, which happens a lot in medicine and small sample studies. The power of meta-analysis is that the summary of these results is quantitative and can be used in the policy-making process. Another contribution of meta-analysis lies in the "prediction" of some outcome under a given scenario (values of variables that significantly impact the effect size). This is based on recreating the aggregate based on the "best-practice" that is also performed in Chapters 3 and 4 in this thesis. A nice example of meta-analysis which was used in the policy debate in the international environment is "Does daylight saving save electricity? A meta-analysis" by Havranek *et al.* (2018), and Chapter 3 of this thesis "The Euro Effect: A Meta-analysis" analyses also a policy-oriented topic.

1.2 Mea-culpa

The referees and the pre-defense of this thesis brought several valuable points that deserve to be clarified, justified, and explained. I decided to explain myself instead of rewriting my research that has been published already. Most of the raised questions are about Chapter 2, "Productivity paradox: A meta-analysis". I started writing that paper 7 years ago, and I used the methods and approaches that I thought were right, which means I found them in papers of the leading academics in the meta-analysis, whom I had the honor to meet at conferences. Seeing my paper now, after almost 10 years, I know I would do many things differently, and one can see it in my latest research: "How bad are trade wars? Evidence from trade costs". In the following paragraphs, I would like to describe the change in my perspective. I would also very much agree with Prof. W. Robert Reed, Ph.D. who wrote, that: "in meta-analysis... in many cases, there is not a clear "correct" way to do things."

My research is also affected by my working experience and experience from supervising bachelor and master theses. I find it more valuable to focus on the topic and data rather than methods. It might be because I perceive my strengths to be in applied research rather than theory – I create more added value by analyzing data than by creating new methods. I put importance on passing on the message extracted from the data and do not let the methods confuse the reader. Therefore when justifying a method I selected, I focus on the positive examples and do not try to find flaws in those. When a new method in empirical research is used and justified, I will gladly adopt it. It does not mean that I will not inform myself about the limitations, but I do not see any added value in repeating them. The following chapters of this dissertation have their added value in the application of meta-analytic methods on original datasets; they do not focus on discussion of the best or the most suitable methods. I realized that using a simpler but suitable method makes it often easier to explain it to others. I would justify it also, in my opinion, that the outcome is crucially dependent on the quality of the input. Next to it, I find it is beneficial to focus on what is positive.

Having said all the above, I would like to address a few issues that meta-analyst faces and how I decided to tackle them along with the explanation. For more details about the best approach, see the "Reporting guidelines for meta-analysis in economics" by Havránek *et al.* (2020).

Outliers. As more data points are collected, the problem with outliers gets more complicated. If the values are strange, one possible approach is to exclude them from the analysis (I did so in Chapter 3). But with more datapoints the distribution of those becomes more heavy-tailed, and one starts to hesitate what to do. I found winsorizing very helpful (applied in Chapter 4). Such a method is commonly used in the applied econometric as, again, it is a quantitative and transparent method with the advantage that observations are not fully thrown away.

Weighted and non-weighted regressions. The first two papers use mostly weighted least squares (WLS) are promoted by e.g., Stanley (2005), while the third paper does not use weighting by standard error as baseline methodology following the up-to-date works like Havranek *et al.* (2017). Reason? WLS is very efficient but has one crucial assumption. The standard errors in the primary studies have to be estimated correctly. WLS method puts more emphasis on the most precise estimates. From the theoretical perspective, that is perfectly ok, because the precision is crucial – the more precise estimate, the more valuable for us. Yet, if one is not sure that the standard error is estimated correctly, another way of weighting is also possible, or one can use it as a robustness check and as baseline methodology regression without weights is used. While the coefficients' interpretation is the same for both cases, non-weighted regression specification makes it easier to explain to the reader and requires fewer assumptions to be fulfilled.

Estimation method. Many possible methods can be used in the meta-analysis, and new methods and approaches are regularly introduced (now the focus is on non-linear methods). My first paper uses a multilevel mixed effects method described by Nelson & Kennedy (2009) in detail and is analogous to the random-effects model widely used in panel-data econometrics. I perceived it as the top-notch methodology, but I realized that it is too specialized for meta-analysis and while estimating it, one does not have full control over what is exactly done, and the replication in other software is more complicated. Given the nature of the data, I find it easier to apply regular panel data estimation methods, especially since those are also available for model-averaging. That is the reason why I now use the OLS and panel data fixed effects. The great benefit is, that these methods are widely known, which means that everybody understands the assumptions, and they are suitable for the data in meta-analysis.

Model selection. There are many different approaches to model and variable selection, starting with no selection at all (my approach in Chapter 2), over the general-to-specific approach, which is based on the elimination of the least significant variables, but such a process is not very robust. In the case of a large number of potential explanatory variables, model uncertainty is a big issue. In the meta-analysis, one of the pioneering work using the model averaging procedure is e.g.Havranek & Rusnak (2013). Since then, the use of Bayesian model averaging (BMA) has gained its space and proved as a very helpful tool and even became part of guidelines for new meta-analysis Economics (Havránek *et al.*, 2020). Similar to other meta-analysts, I use BMA in the second and third papers. BMA provides a weighted average of individual regressions that each include a different combination of explanatory variables. The weights are derived from the posterior model probabilities, putting larger weight to more probable results. BMA does not estimate all potential models since the computing capacity is not sufficient yet, but allows to reduce the model space to deal with the most probable models. Even for model averaging, one can specify priors (ex-ante limits and restrictions), which affect the outcome but help to limit the model space. Therefore, it is desirable to either use more general conditions or provide robustness checks with different priors. Model averaging ensures that the final model specification is not handpicked (priors only restrict the model space and do not select one and only specification), and hence results are more robust compared to other more volatile approaches. The disadvantage of model averaging is, that when using pre-defined software packages, one cannot fully control the conditions used for the model selection (e.g. if some assumptions hold). Havranek (2019) nicely summarizes why model averaging should be part of the meta-analysis.

1.3 This dissertation chapters in more detail

The first paper of this dissertation is named "Productivity paradox: A meta-analysis" and focuses on the ICT investments and their effect on productivity. The productivity paradox has been investigated for decades, and from a large amount of published literature, one can hardly expect to make a general conclusion without using proper summarizing techniques. The existing literature has focused on finding a positive effect to refute the paradox. I employed the meta-analysis using more than 850 estimates from more than 70 studies and found that such evidence suffers from substantial publication bias. The estimate of the underlying effect beyond the publication bias is about 0.3%, which is about ten times lower than usually expected. Since the size of the effect helps make the right decision in business-related investments, our result of ICT elasticity being very close to zero for productivity and no effect (not significantly different from zero) on profitability supports the argument that there are better forms of investment to be made.

The second paper of this dissertation is named "The Euro Effect: A Meta-analysis" and focuses on the effect of monetary unions and especially on the eurozone. In the area of international economics, bilateral trade flows are analyzed very profoundly, and it is expected that common currency should support the trade (Pugh *et al.*, 1999). Such effect has been found on several currency unions, and the most known work in that area is of Rose & Stanley (2005); therefore, it is also often labeled as the "Rose effect". Rose estimated the effect to be up to 200%, which is enormous. A few years later, Havranek (2010) made a meta-analysis to investigate the Rose effect among existing literature and found serious publication bias and no effect at all. The second paper of this dissertation follows Havranek (2010) and collects over 3300 estimates from almost 60 studies, which makes it the largest meta-analysis in terms of estimates[1]. I confirmed the presence of publication bias and estimated the euro's effect on common trade to be around 2%. That is much less than the original Rose effect, but it is still positive. The possible explanation is that the euro-area is part of the European Union (EU), which already removes trade barriers and promotes cooperation. Therefore, the benefits of a single currency on trade are minimal, especially if the financial markets are entirely integrated and efficient. The only costs the trading parties face are exchange risk, but it

[1] I collected 28 characteristics for the heterogeneity estimation, but today meta-analysts collect even 70. Therefore it is not the largest if the measure is total datapoints.

can be easily hedged on the financial markets and menu costs. The paper itself is not supposed to argue pro or against the euro currency but provides sound evidence about its effect on one single part of the economy – trade.

The third paper of this dissertation is named "How bad are trade wars? Evidence from trade costs" and focuses again on international economics and bilateral trade flows as the second paper. However, this time, the paper is co-authored by Nikol Poláková and Anna Tlustá and also provides the theoretical perspective about the gravity equation that is used widely. The paper focuses mainly on trade costs in terms of tariffs and their effect on the bilateral trade flows. Tariffs are creating trade barriers, and trade agreements are trying to lower them to promote trade and welfare. However, countries all over the world use tariffs as protectionist measures or as a policy tool. When writing these lines at the beginning of 2020, the world's largest economies – the USA and China – are discussing whom to lower the tariffs after raising them before. Great Britain left the EU and now has to set up the new rules, including the new trade conditions and tariffs. The results presented in the paper show again a strong presence of publication bias and the effect beyond it to be very small. Yet it is statistically different from zero, which clearly supports the theoretical concepts formulated decades ago. Trade wars are bad and reduce trade and welfare.

Bibliography

ALINAGHI, N. & W. R. REED (2018): "Meta-analysis and publication bias: How well does the fat-pet-peese procedure work?" *Research synthesis methods* **9(2)**: pp. 285–311.

ASHENFELTER, O. & M. GREENSTONE (2004): "Estimating the value of a statistical life: The importance of omitted variables and publication bias." *The American Economic Review* **94(2)**: pp. 454–460.

BURMAN, L. E., W. R. REED, & J. ALM (2010): "A call for replication studies." *Public Finance Review* **38(6)**: pp. 787–793.

CARD, D. & A. B. KRUEGER (1995): "Time-series minimum-wage studies: A meta-analysis." *The American Economic Review* **85(2)**: pp. 238–243.

CHRISTENSEN, G. & E. MIGUEL (2018): "Transparency, reproducibility, and the credibility of economics research." *Journal of Economic Literature* **56(3)**: pp. 920–80.

CHRISTENSEN, P. (2003): "Topics in meta-analysis. a literature survey." *Technical report*, Institute of Transport Economics, Oslo.

DEWALD, W. G., J. G. THURSBY, & R. G. ANDERSON (1986): "Replication in empirical economics: The journal of money, credit and banking project." *The American Economic Review* **76(4)**: pp. 587–603.

DOUCOULIAGOS, C. & T. D. STANLEY (2008): "Theory competition and selectivity: Are all economic facts greatly exaggerated?" *Technical Report 6.*

DUVENDACK, M., R. PALMER-JONES, & W. R. REED (2017): "What is meant by" replication" and why does it encounter resistance in economics?" *American Economic Review* **107(5)**: pp. 46–51.

DUVENDACK, M., R. W. PALMER-JONES, W. R. REED *et al.* (2015): "Replications in economics: A progress report." *Econ Journal Watch* **12(2)**: pp. 164–191.

GLASS, G. V. (1976): "Primary, secondary, and meta-analysis of research." *Educational Researcher* **5(10)**: pp. 3–8.

HAVRANEK, T. (2010): "Rose effect and the euro: Is the magic gone?" *Review of World Economics* **146(2)**: pp. 241–261.

HAVRANEK, T. (2019): "Why model averaging is useful in meta-analysis."

HAVRANEK, T., D. HERMAN, & Z. IRSOVA (2018): "Does daylight saving save electricity? a meta-analysis." *The Energy Journal* **39(2)**.

HAVRANEK, T. & M. RUSNAK (2013): "Transmission lags of monetary policy: A meta-analysis." *International Journal of Central Banking* **9(4)**: pp. 39–76.

HAVRANEK, T., M. RUSNAK, & A. SOKOLOVA (2017): "Habit formation in consumption: A meta-analysis." *European Economic Review* **95**: pp. 142–167.

HAVRÁNEK, T., T. STANLEY, H. DOUCOULIAGOS, P. BOM, J. GEYER-KLINGEBERG, I. IWASAKI, W. R. REED, K. ROST, & R. VAN AERT (2020): "Reporting guidelines for meta-analysis in economics." *Journal of Economic Surveys* .

HUNTER, J. & F. SCHMIDT (2004): *Methods of meta-analysis: correcting error and bias in research findings*. Sage, 2 edition.

IOANNIDIS, J., T. STANLEY, & C. DOUCOULIAGOS (2017): "The power of bias in economics research." *Economic Journal* **127(605)**: pp. F236–F265.

NECKER, S. (2014): "Scientific misbehavior in economics." *Research Policy* **43(10)**: pp. 1747 – 1759.

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(3)**: pp. 345–377.

NORMAND, S.-L. T. (1999): "Meta-analysis: formulating, evaluating, combining, and reporting." *Statist. Med.* **18(3)**: pp. 321–359.

PETITTI, D. B. (2001): "Approaches to heterogeneity in meta-analysis." *Statist. Med.* **20(23)**: pp. 3625–3633.

PUGH, G., D. TYRRALL, & L. TARNAWA (1999): "Exchange rate variability, international trade and the single currency debate: a survey." *Economic Policy in the European Union: Current Perspectives* pp. 9–25.

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.

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. & H. DOUCOULIAGOS (2019): "Practical significance, meta-analysis and the credibility of economics." .

Chapter 2

The Productivity Paradox: A Meta-Analysis

Abstract[1]

The impact of information and communication technology (ICT) on economic performance has been the subject of academic research for several decades, and despite the remarkable and significant innovation in computer technology, usage, and investments, only a small growth in productivity has been observed. This observations has been coined the productivity paradox. This paper uses meta-analytical methods to examine publication bias and the size of the output elasticity of ICT capital. The empirical part is based on a collection of more than 800 estimates of ICT payoff effects from more than 70 studies written in the last 20 years. The meta-analysis reveals a strong presence of publication bias within the ICT productivity literature and, using a mixed effect multilevel model, estimates the output ICT elasticity to be only 0.3%, which is more than ten times smaller than what was reported by a previous meta-analysis from 10 years ago.

[1] The paper was published in the *Information Economics and Policy* [2017, 38, pp. 38—54]. I gratefully acknowledge helpful suggestions by the editor and anonymous referees. I thank participants of MAER-Net Conference in Athens 2014 for helpful comments. I acknowledge support from the Grant Agency of Charles University (grant #280515). The views expressed in this paper are those of the author and not necessarily those of the Czech National Bank.

2.1 Introduction

The economic significance of productivity is well established, and productivity as such is an indivisible part of economic theory. Productivity not only determines wealth and economic growth as well as being an indicator of competitiveness but also creates foundations for management decisions not only at the firm level but also at the national level for policy makers. Productivity growth is a desirable outcome of development and technological progress. Since information and communication technology (ICT) gained in importance in production processes in the second half of the twentieth century, especially with the emergence of the so-called "new economy" (service-based economy), a research field focusing on the relationship between ICT and productivity has emerged. This is also the aim of this paper.

Economics tries to find the most effective allocation of resources, and ICT is more and more embedded into the production process. In the 1980s, studies found no evidence of increased productivity due to ICT investments (Mahmood & Mann, 1993). In Solow (1987), Robert M. Solow, a Nobel laureate in economics, wrote in a book review that "You can see the computer age everywhere but in the productivity statistics". His famous quip may have been the starting point of decades of research and discussions on the influence of ICT on productivity. Economists such as Brynjolfsson (1993); Harris (1994); Willcocks & Lester (1996); Brynjolfsson & Hitt (1998) point to this phenomenon as the "productivity paradox".

The productivity paradox opened up a wide debate and the continuous examination of the topic and also attempts to explain the theory contradicting results (Dedrick *et al.*, 2003; Stiroh, 2005; Draca *et al.*, 2006). Dedrick *et al.* (2003) provided a narrative review of the published research and evidence from more than 50 articles refuting the productivity paradox. Kohli & Devaraj (2003) were the first to summarize empirical results at the firm level in a meta-analysis and clearly stated that there is no productivity paradox. Later, Stiroh (2005) in his meta-analysis included studies based on more aggregated data and mentioned possible publication bias among the results but did not take any steps to check for such bias. After almost a decade, extension of Stiroh (2005) provided by the narrative review Cardona *et al.* (2013) brings more insight into the development of the effect size over time in relation to different approaches and by examining ICT being general purpose technology (GPT) following David (1990) but does not aim to conclude anything about the size of the effect.

To analyse the available empirical results in a quantitative way, we used meta-analytic methods in this paper. Meta-analysis also makes it possible to test for the presence of publication bias; therefore, we can test the hypothesis made by Stiroh (2005) that such bias exists in the productivity paradox oriented literature. Furthermore, meta-analysis allows us to estimate the true size of the underlying effect even if publication bias is present. This paper follows Dedrick *et al.* (2003); Kohli & Devaraj (2003); Stiroh (2005); Stanley (2008); Cardona *et al.* (2013), uses the meta-analytical techniques to test for publication bias in the productivity paradox literature, and attempts to explain how different study characteristics influence the size of the estimate of the effect size of ICT investments on productivity.

Compared to the previous studies in the area of meta-analysis and IT productivity payoff performed by Kohli & Devaraj (2003); Stiroh (2005); Cardona *et al.* (2013), there are several aspects that make this paper unique. First, a publication bias treatment has not yet been done. Second, never have all the estimates from each study been collected, so a modern econometric framework that takes study heterogeneity into account has not been used in this area of research until now. Third, this paper estimates the effect size of the ICT payoff using all available research, whereas the previous research was focused on factors that might influence the effect size. Fourth, this study is the first one to estimate the effect at the firm level using explanatory meta-analysis regression.

This paper is structured as follows. Section 2.2 provides insight into production process modelling, followed by section 2.3, which summarizes the main empirical findings of the productivity paradox. Section 2.4 is devoted to meta-analysis, an econometric technique that is used in the empirical part. Section 2.5 provides descriptions of the data, followed by section 2.6, where the results from the conducted analysis are highlighted. The final notes conclude.

2.2 Production process modeling

To analyse the contributions of the production factors to the output, it is necessary to describe the production process mathematically and to create an economic model. We can simply imagine the process as a functional relation between outputs and inputs. Therefore, the most used approach by economists to the model production process is using specific production functions that algebraically formulate the relation between the inputs and the outputs. A simple version of an aggregate production function that puts together similar inputs was described by Solow (1957) in a form that is depicted in equation 2.1. Q represents output; K capital; L labour; A is a multiplicative factor that captures the technological development, which determines how efficiently the inputs are used to produce output; and $f(\cdot)$ represents the functional relation.

$$Q = Af(L, K) \tag{2.1}$$

The standard approach in the economics literature is to consider the Cobb-Douglass production function (Brynjolfsson & Hitt, 2000; Stiroh, 2005; Venturini, 2009), which can be expressed as

$$Q = AK^{\alpha}L^{\beta} \quad \alpha, \beta > 0 \tag{2.2}$$

For econometric estimation, the logarithmic form (eq. 2.3) of the Cobb-Douglas production relation is more useful. The coefficients of interest are α and β, which denote the elasticities of capital and labour, respectively.

$$\ln Q = \ln A + \alpha \ln K + \beta \ln L \quad \alpha, \beta > 0 \tag{2.3}$$

To estimate the effect of ICT, we need to separate capital into ICT (K_{ICT}) and non ICT (K_{ICT}). Next, studies may vary in additional firm or industry specific inputs (M) that are part of the production. The simplest form of the production function used for estimation that includes the abovementioned 2 types of capital, labour and intermediate inputs is depicted in equation 2.4, where ϵ stands for the disturbance term.

$$\ln Q = \alpha + \beta_{IT} \ln K_{IT} + \beta_{nonIT} \ln K_{nonIT} + \beta_L \ln L + \beta_M \ln M + \epsilon \tag{2.4}$$

In general, using model 2.4 augmented by several inputs, we are able to determine how each factor influences the production. Effects that are measured by the presented framework are mostly direct, but since ICT costs are by nature investments, indirect effects (spillovers) are also present. Evidence about indirect effects are provided in, e.g., Mittal & Nault (2009) or Han *et al.* (2011); however, in our analysis, we focus on direct effects, which are important to the decision making of individual firms.

2.3 Productivity paradox

Productivity growth arises from innovation and development of new production methods, procedures and technology. Therefore, ICT as a new technology was expected to boost the productivity. However, the initial results were inconclusive (Barua *et al.*, 1995; Teo *et al.*, 2000), and hardly any positive effects of ICT investments on productivity were found. This paradox is called the *productivity paradox*, more precisely and formally defined by Turban *et al.* (2002, 592) as "The discrepancy between measures of investment in information technology and measures of output at the national level." Methods and models used to analyse the quantitative data are mostly based on the neoclassical production theory that clearly predicts the sign and magnitude of capital's elasticity: in the case of constant returns and competitive markets, elasticity should be equal to the share of the factor. Initially, however, this relationship was not found in the data.

The first pieces of the economics literature trying to clarify the paradox were of a narrative and descriptive nature. Then, studies such as Thatcher & Oliver (2001) employed microeconomic theories and models. In addition, theoretical estimations of the elasticity of ICT investments using simulated data were conducted by, e.g., Yorukoglu (1998). Narrative explanations provided first by Baily *et al.* (1988); David (1990); Brynjolfsson (1993) and later by Triplett (1999); David (2000); Horzella (2005); Cardona *et al.* (2013); Biagi *et al.* (2013) can be divided into three general categories, or perspectives, depending on how they approach the productivity paradox: 1) measurement problems, 2) context and mismanagement, and 3) ICT as GPT.

David (1990) put the productivity paradox into historical perspective by comparing computers to steam engines and electricity. These inventions needed decades until their contributions became visible, and the same holds for computers, which are the backbone of ICT today. Such technology is called GPT. Brynjolfsson (1993) formulated and discussed four possible reasons why the empirical literature failed to find positive returns on ICT investments. Later, Triplett (1999) discussed seven possible explanations for Solow's productivity paradox and empirical evidence dismissing the paradox, such as Oliner & Sichel (2000), followed.

Because of the huge research interest in ICT payoff, empirical studies summarizing ICT productivity also emerged, as discussed in the introduction. Kohli & Devaraj (2003), unlike a standard meta-analysis, do not examine the effect size but only the research factors that contribute to discovering a relation between ICT and firm performance. The authors use a wide range of literature reports and use data from various research approaches and models. Therefore, only the signs of the explanatory variables with any effect can be interpreted, not the magnitude. In addition, they mainly compare studies reporting positive and negative estimates regardless of the magnitude and precision of the effect. Stiroh (2005) estimates the underlying true effect – the elasticity of ICT investments – based on the previous results. He calculates the pure arithmetic mean value from 20 estimates in the range between -0.06 and 0.24 and obtains a result equal to 0.054; in addition, using fixed effects and the OLS method, he finds a significant estimate of 0.065. He was followed by Cardona *et al.* (2013), who extend the examined literature and put the effect size on the time line but do not provide any conclusions about the effect size. Biagi *et al.* (2013) reviews also literature with alternative approaches to measure the effect of ICT and its derivatives as information systems, product and process innovation, cooperation, management etc. Biagi *et al.* (2013) concludes that evidence confirms ICT is a GPT.

None of the previous empirical review studies take the precision of the estimate into account. Stiroh (2005); Cardona *et al.* (2013) use both firm level and more aggregated level (industry and national) results as inputs, which is not performed by Kohli & Devaraj (2003) nor by this paper. Since we discuss ICT payoff and meta-analysis research, we would also like to mention a related study performed by Lim *et al.* (2011). The focus of the authors is the effect of ICT on firms' financial performances using only one estimate per paper, and they do not use any regression techniques. This paper fills in the gaps from the previous research.

2.4 Meta-analysis methodology and publication bias

The basic idea of meta-analysis is to examine the factors that influence the research results of some phenomena. The dependent variable is the effect size of each estimate, while the independent variables consist of various information about each study such as the data characteristics, method used for analysis, sample size, and sometimes even the occupation of the researcher. Glass (1976, p.1) defined "meta-analysis" as the statistical "analysis of analyses", a tool for integrating findings from a collection of individual studies. Detailed overviews of contemporary methods used in meta-analysis can be found in, e.g., Nelson & Kennedy (2009) or Stanley *et al.* (2008), which are also sources for this paper.

Studies provide estimates of different sizes, and in cases when the variance of the results is too large to be justified by the disturbance terms, we speak about the heterogeneity between studies. We try to explain this heterogeneity by specific differences between the studies. Thus, we code the properties of the studies into variables, and later, we test for the presence of heterogeneity. As Christensen (2003) describes, there are two general types of heterogeneity present in the research: factual and methodological. Factual heterogeneity concerns real differences in the effect due to actual differences in the tested sample, for example, when a study was conducted at a different time or in a different country. In our case of ICT capital, there could be a difference between developing countries, where the economies are based on manufacturing, and developed countries, where the economies are service oriented. Methodological heterogeneity is the result of different study approaches, which could be the models used, the data characteristics, or the econometric methods. Christensen (2003); Nelson & Kennedy (2009) present two common ways of how to deal with the heterogeneity in the data. The first one uses moderating variables and meta-regression to detect the sources of the heterogeneity, the second approach uses random effect size (RES) models. This paper uses both methods, but the RES model is replaced with a more advanced mixed effects model.

While determining which study characteristics influence the results, we might overlook an important factor that might cause significant bias of the outcome – the researcher.Sterne *et al.* (2000) argue that published results are biased due to publication or selection bias. Such bias stems from researchers' motivation to get their work published, i.e., to provide results that are unlikely to be rejected by journals' reviewers. If the outcome of an empirical study is not in line with the underlying economic theory, then something is wrong, and reviewers would require proper explanation. It is easier to select a model or data that fits the expectations. Stanley (2005) has two categories for publication bias. Any modification (e.g., restraining the dataset or model modification) to obtain results that are in line with the theory is labelled as publication bias of *type I*. Adjusting of models may also happen in case the standard procedure has an insignificant outcome, and it is not the desired outcome. *Type II* publication bias can be therefore described as reaching statistical significance no matter what the effect size would be.

Doucouliagos & Stanley (2011) focused on publication bias, gathered and analysed several thousand estimates from approximately three and half thousand separate empirical studies from 87 different areas of empirical economic research (the productivity paradox was not included). A large portion of these fields has been found to be burdened by publication bias. In particular, macroeconomic research contains evidence of severe selectivity that significantly distorts empirical findings. This only confirmed that publication bias has been a serious issue in empirical economics research as argued by Long & Lang (1992); Card & Krueger (1995); Ashenfelter & Greenstone (2004); Stanley (2005), and it might be present in the productivity paradox literature as well.

Stiroh (2005) said that the "evidence clearly points to a positive productivity effect from IT", but we could compare this to the area of the European monetary union and trade benefits. The so-called Rose-effect also clearly pointed to a positive effect of the European monetary union (EMU) (Rose, 2000), but Havranek (2010) found no significant return of the EMU after he took publication bias into account. The theory of currency unions expects that after integration, common trade will increase, and this argument is widely used in politics to promote this type of integration. Havranek's (2010) examination discovered a skewed distribution of the estimates and subsequently no significant genuine effect. For other monetary unions, the effect was found to be positive and significant. A possible explanation is that the European Union creates such close integration that there are no significant gains from the common currency. The same might be the case of ICT, with the studies in the 1980s being correct, and the only positive effect we can see in the literature is caused by the publication bias.

2.4.1 Meta-regression model

Publication bias is usually detected using two methods – graphical and quantitative. The first one is an informal examination, but in some cases (e.g. publication bias in effect of currency unions discovered by Rose & Stanley (2005) and confirmed by Havranek (2010)), it provides sufficient evidence. Econometric methods can not only discover the presence of the publication bias but also estimate the true effect beyond (Hunter & Schmidt, 2004). In the empirical part, the 2.8 and 2.9 models are used. The logic behind those models is explained step by step in the following part.

The basis for publication bias detection is formed by the funnel plot, also called the funnel diagram, which is widely used in meta-analyses mainly as a tool for detecting possible publication bias (Egger *et al.*, 1997; Sterne *et al.*, 2001). Individual observations (estimates of effect sizes) are plotted on the horizontal axis against a measure of the precision on the vertical axis, mostly inverted standard error or the square root of sample size. Large studies will show lower variation then small studies that are less precise. This should generate a plot that looks like an inverted funnel with the most precise estimates (with the shortest confidence intervals) on the top and less precise estimates on the bottom. The theory of funnel symmetry originates in the idea that there is one underlying population value – it can even be zero – and a probability distribution that would converge to a normal distribution with the mean value equal to the underlying true effect. In the case of funnel asymmetry, especially when the plot is skewed or one side is missing, we should be suspicious of publication bias. On the other hand, the apparent symmetry might not foreclose the publication bias. Extensive description of funnel plots is provided by Stanley & Doucouliagos (2010).

Econometric testing for publication bias follows this logic; thus, if reported estimates are dependent on their standard errors (Card & Krueger, 1995),

$$b_i = \beta + \alpha_0 \cdot se_i + u_i, \quad u_i \sim N\left(0, \delta^2\right) \tag{2.5}$$

In the model 2.5, the estimate b_i depends on its standard error (se_i). The degree of dependence is measured by the coefficient α_0, which represents the degree of the publication bias, and if it is significant, we have a formal proof for funnel asymmetry. However, model 2.5 suffers from heteroskedastic se_i. To solve this issue, we follow recommendations of Stanley *et al.* (2008), who suggests the usage of weighted least squares (WLS) in the form of where standard errors are used as weights, which will result in the dependent variable being the t-statistic:

$$\frac{b_i}{se_i} = t_i = \frac{\beta}{se_i} + \alpha_0 + \xi_i, \quad \xi_i \sim N\left(0, \sigma^2\right) \tag{2.6}$$

There is also an issue of within study heterogeneity. Studies usually present more than one estimate of the effect size. Estimates thus share the same dataset and methods and are likely to be highly correlated. As a result, we cannot handle them as independent values. This issue has been known for a long time (Stanley & Jarrell, 1989). Examples and discussions how to solve possible dependence can be found in Johnston *et al.* (2003); Bateman & Jones (2003); Bickel (2007); Gelman & Hill (2006); Hox (1995); Hox & Leeuw (2003); Peters *et al.* (2010), and the most commonly used remedy is described by Nelson & Kennedy (2009); Doucouliagos & Laroche (2009) and called the mixed-effects multilevel model. As Nelson & Kennedy (2009) further elaborate, the mixed-effects multilevel model is analogous to the random-effects model that is widely used in panel-data econometrics. The mixed-effect model is a combination of models with fixed effects (β) and a random part (ζ_j) that gives the flexibility to the model and is therefore better for meta-analytic purposes. It considers the diversity of the data and also allows multiple random effects. By extending model 2.6, we obtain a model in the same way as Havranek & Irsova (2011):

$$t_{ij} = \frac{\beta}{se_{ij}} + \alpha_0 + \zeta_j + \epsilon_{ij}, \quad \zeta_j \sim N(0, \psi), \quad \epsilon_{ij} \sim N(0, \theta) \tag{2.7}$$

In the final specification of model (2.7), j and i denote the index of the study and the estimate within the study. The t-statistics are therefore analogically labelled (t_{ij}). As we have j studies, we have in total $J = \sum J_j$ estimates. The overall error term (ξ_{ij}) consists of study-level random effects (ζ_j) and estimate disturbances (ϵ_{ij}). We assume independence of both values; thus, we can simply sum their variances to find the composite error: $\text{Var}(\xi_{ij}) = \psi + \theta$ with θ being the within-study variance and ψ the between study variance. The closer the value of ψ is to zero, the less advantageous the usage of the mixed-effects framework is instead of OLS.

Since the aim of this paper is not only to determine the effects of ICT investments but also to explain the variation of the reported values, we use the characteristics of individual studies and, following Stanley & Jarrell (1989); Stanley *et al.* (2008), add a vector of explanatory variables Z_k to model 2.7:

$$t_{ij} = \frac{\beta}{se_{ij}} + \alpha_0 + \sum_{k=1}^{K} \frac{\gamma_k Z_{kij}}{sei_{ij}} + \zeta_j + \epsilon_{ij}, \quad \zeta_i \sim N(0, \psi), \quad \epsilon_{ij} \sim N(0, \theta) \tag{2.8}$$

The explained variable is the t-statistic and not the estimate of the effect size, explanatory variables Z_k have to be also weighted by standard error.

As we are also attempting to find the magnitude of the ICT payoff "cleaned" from the publication bias, we follow Stanley & Doucouliagos (2007); Havranek & Irsova (2016); Stanley & Doucouliagos (2014) and augment model 2.7 with an additional standard error variable (which stems from the possibility that the standard errors effect can be quadratic), which results in the so-called Heckman meta-regression.

$$t_{ij} = \frac{\beta}{se_{ij}} + \alpha_0 \cdot se_{ij} + \zeta_j + \epsilon_{ij}, \quad \zeta_j \sim N(0, \psi), \quad \epsilon_{ij} \sim N(0, \theta) \tag{2.9}$$

where β reports the magnitude of the underlying effect corrected for the publication bias. This last specification completes the framework needed for our analysis. The dataset used for this task is described in the next part.

2.5 Dataset

No dataset exists for our purposes, and thus, all data had to be retrieved manually. The base for the list of studies was provided by the previous works mentioned in the introduction, namely Kohli & Devaraj (2003); Stiroh (2005); Cardona *et al.* (2013). Collecting all of the studies used by Stiroh (2005); Cardona *et al.* (2013) was completed without any issues, but the retrieval of the studies included in the second meta-analysis by Kohli & Devaraj (2003) was more problematic because we were not able to obtain the complete list of studies due to limited availability.[2] Since researches are still investigating the effects of ICT on productivity, the literature sample was further extended by searching on the RePEc website. This database not only covers all journals[3] used by the two previous studies but also extends the searching area.

The search criteria followed previous empirical studies only, which included papers written after 1990. Therefore, the search was focused on "productivity paradox", "ICT productivity" and "information technology"[4], and only items with empirical firm level research written in English since 1990 were taken into account. This search restriction makes this study comparable to the previous research.

For proper meta-analysis, it is important to have a coherent research design so we can compare the results, especially when we focus on publication

[2] We were not able to retrieve a few dissertations, which we could not find online, and the authors did not respond to our requests, namely – Alshilash (1997); Cline (1999); Chen (1997); Mayberry-Stewart (1996).

[3] With the exception of *Information Systems Research*, *Information & Management* and *Journal of Management Information Systems* that were searched using `isr.journal.informs.org`

[4] The exact search query used in RePEc was firm + ((information + technology + (productivity — payoff)) — (productivity + paradox) — (ICT + (productivity — payoff))) + estimate and also term ((information+technology)—ICT)+ investment + firm + productivity, searched in abstract since 1990.

bias. We checked whether all studies use the production function framework as introduced in the second chapter, and work with ICT capital as a separate explanatory variable, thus we excluded papers testing the probability of, e.g., product innovation via probit or logit models. Next, we only included studies that somehow calculated the value of the ICT capital – therefore, we did not take into account studies using only dummy variables for usage or non-usage of ICT, e.g., Atrostic & Nguyen (2005). Lastly, only studies using firm-level data were selected. All these restrictions are needed for the studies to be comparable and suitable for an aggregated analysis, which is also the reason why several studies that Kohli & Devaraj (2003) and Stiroh (2005) used were excluded due to an inconsistent framework.

We need a framework where we can observe ICT elasticity, i.e., log-log or translog models, as we cannot compare coefficient estimates to, e.g., level-log models. Let us recall equation 2.4:

$$\ln Q = \alpha + \beta_{IT} \ln K_{IT} + \beta_{nonIT} \ln K_{nonIT} + \beta_L \ln L + \beta_M \ln M + \epsilon$$

For us, the important coefficient is β_{IT}, which is the output elasticity of ICT capital (further also labeled as elasticity of ICT capital). In other words, it tells us that the increase in K_{IT} by 1 percent increases Q by β_{IT} percent. We focus on ICT investments related to productivity or profitability. Therefore, studies analysing, e.g., the effect of ICT on the technological progress represented by some index could not be used. Neither studies where the target is to lower something, e.g., mortality (Devaraj & Kohli, 2003), nor measure the scale of something, e.g., process changes (Grover *et al.*, 1998) or production hours (Kelley, 1994), can be used. Additionally, studies using pure growth accounting approaches or correlations only and therefore no regression, such as (Kivijärvi & Saarinen, 1995; Lubbe *et al.*, 1995), were excluded because we do not have the precision indicators for such studies.

In total, we identified 55 published studies and 16 working papers. These 71 works written between 1992 and 2016 report more than 800 estimates of ICT elasticity. To get an idea of the sizes of those numbers, we can compare them to the numbers from Nelson & Kennedy (2009), who report the mean and median of 191 and 92 observations; the mean and median of these studies are equal to 42 and 33, respectively. All those values are based on a survey consisting of 125 meta-analyses. Our sample greatly exceeds both indicators. The list of the studies together with the number of estimates and indication if they were previously used in one of the mentioned meta-analysis is presented in table 2.9.

After suitable studies were identified, gathering and coding of all variables was carried out. The most important variables were of course the estimate of ICT elasticity and respective t-statistic, standard error or significance level only when reported. Various measures of productivity can be used at the firm level; thus, we created two groups of dependent variables and coded the measures according to it. The first group aggregates productivity measures, such as output or sales, and the second group is related to the profitability and thus mostly financial measures such as return on assets (ROA) or return on equity(ROE). In most cases, the industry type was not reported or industries were mixed. Thus, dummies for services and production types of industry were not used.

As the main target of this paper is to determine publication related effects, variables that are likely to have influences on the effect size were collected, i.e., the econometric *method* used for estimation, *sample size*, *data source*, *time span* and *average year* of the data in the sample. When exploring the publication bias, we are not only interested in when and if the work was published or not but also how the study was further used. As a proxy, we use the *number of citations* of the study provided by Google Scholar in early August 2016. Using the absolute number of citations would handicap newly published works; thus, we normalize the number of citations using the year of publication and taking the logarithm, as is the standard approach. Journals and other *places of publication* were also coded, which will allow us use the multi-level mixed-effects model. Regional differences are captured in region dummies – we created 2 main regions based on the data source: US and the rest of the world. Table 2.7 provides a list of the variables used.

2.5.1 Data description

We collected a nearly complete data matrix: the number of observations in the studies varies from 24 to 36305 (Tambe & Hitt, 2012) and an average study uses 3145 observations for estimates. We gathered the most estimates (113) from Brynjolfsson & Hitt (2003), but half of the studies report less than 9 estimates of the effect size. The average and median publication year are 2003 and 2002, respectively – in the year 2002, the dataset for the last meta-analysis by Stiroh (2005) ends. In total, we identified 67 Journals and working paper sources and

22 sources of data. More than half of the studies (44) have not been used before for meta-analysis of ICT productivity. A further basic dataset description is provided in table 2.1.

Table 2.1: Basic dataset description

	Min	Max	Mean	Median
$\hat{\gamma}$ per study	1	113	12	9
observations in study	24	36305	3145	1148
time span (in years)	1	30	8	6

$\hat{\gamma}$ – estimate of ICT elasticity
numbers are rounded to the nearest integer

A closer examination of the collected estimates is provided in table 2.2. We divided the estimates into groups depending on some of the characteristics related to publication bias. The minimum and maximum values are far from the mean value in all samples. The highest estimates - close to one - are from Rai *et al.* (1996). This study uses the whole information system budget and the production function lacks several explanatory variables. The next highest results come from Zwick (2003), where the anomaly is caused by the MLE method and data selection procedure. The results for the non-selected data are around 0.05 when estimated using the OLS method. The other side of the interval, with largely negative values, is covered by Paton *et al.* (2004) who conduct a case study of the gaming industry.

Table 2.2: Descriptive statistics of the dataset – elasticity estimates

	Mean	Std. Dev.	Min	Max	Average*	Obs
Full sample	0.0643	0.1385	-1.124	0.994	0.0035	850
Kohli	0.0740	0.1985	-.516	0.994	0.0110	133
Stiroh	0.0334	0.0432	-0.086	0.222	0.0033	253
New studies	0.0762	0.1430	-1.124	0.98	0.0203	510
Working paper	0.0523	0.1903	-1.124	0.98	0.0335	157
Published	0.0670	0.1238	-.516	0.994	0.0034	693

*average = weighted average calculated using formula 2.10
Kohli – studies also included in Kohli & Devaraj (2003)
Stiroh – studies also included in Stiroh (2005)
new studies – not used by Kohli or Stiroh

One of the basic meta-analytic techniques is a calculation of weighted averages of all available estimates. Such an approach to the ICT productivity literature was taken by Lim *et al.* (2004), but the aim of this work was to determine the differences between specific groups of firms, not to find the underlying true effect of ICT investments. The calculated weighted average for each group is also included in table 2.2 using formula 2.10. The inverted variances of the estimates are used as weights.

$$r_w = \frac{\sum_{i=1}^{n} w_i \gamma_i}{\sum_{i=1}^{n} w_i}, \quad w_i = \frac{1}{Var(\gamma_i)} \tag{2.10}$$

Weighted averages are much smaller than simple averages, and there is no common ratio for all groups. Thus, the smaller mean in one group does not necessarily result in a smaller weighted average, like for working papers. On average, working papers report smaller effect sizes than published papers, but the weighted average of the published estimates is ten times smaller than the weighted average of the working paper estimates. This is caused by several very precise estimates that are very close to zero since the weight is the inverse of the variance in formula 2.10. As the estimate is closer to zero, it has to be more precise in order to be significant at the generally respected 5% significance level.

There is also a difference between the studies selected by Kohli & Devaraj (2003) and Stiroh (2005). The former one uses studies with higher and more diverse estimates (larger variance). In addition, studies that were not used in either of those papers contain on average higher effect sizes. This may be the time effect or, e.g., due to the selection process. The main conclusion of the basic stratification of the data is that the mean value of the effect size is around 0.06, but the weighted average is close to zero with values almost twenty times smaller, reaching 0.0035. If we use the inverted standard errors as weights, the values are around 0.03.

As for the groups, the weighted average was calculated for each of 72 studies with the weights equal to the inverse of the variance. The result is depicted by the forest graph in figure 2.1, where the dashed line close to zero represents the weighted average of all studies estimated using the fixed effect size FES method (model 2.5), which is equal to 0.004 with z-stat. = 50.50. This is the simplest model used in meta-analysis, but it has some disadvantages as mentioned above. The result is influenced by studies such as Lehr & Lichtenberg (1999) and Hitt & Brynjolfsson (1996), who report significant estimates very close to zero (0.00061 and −0.0008, respectively). The weight of each study in the final result is illustrated by a grey a square in the figure, and the abovementioned studies dominate the others. For the other estimates, the squares, and therefore the weights, are much smaller. If we remove these two studies, we immediately end up with a weighted average equal to 0.029 and z-statistic equal to 134.6. Without any data modification, the basic model with the random effect provides $\hat{\gamma} = 0.036$ and $CI = (0.034; 0.037)$, z-stat. = 49.54. We can clearly see how the different methodology provides us with a completely different result. The reason why we use the multilevel mixed effects model for the final analysis is following: the estimates from each study are mutually correlated because they are based on the same data set and methodology and have to be treated that way. We combine the fixed effect within the study with a random component among the studies.

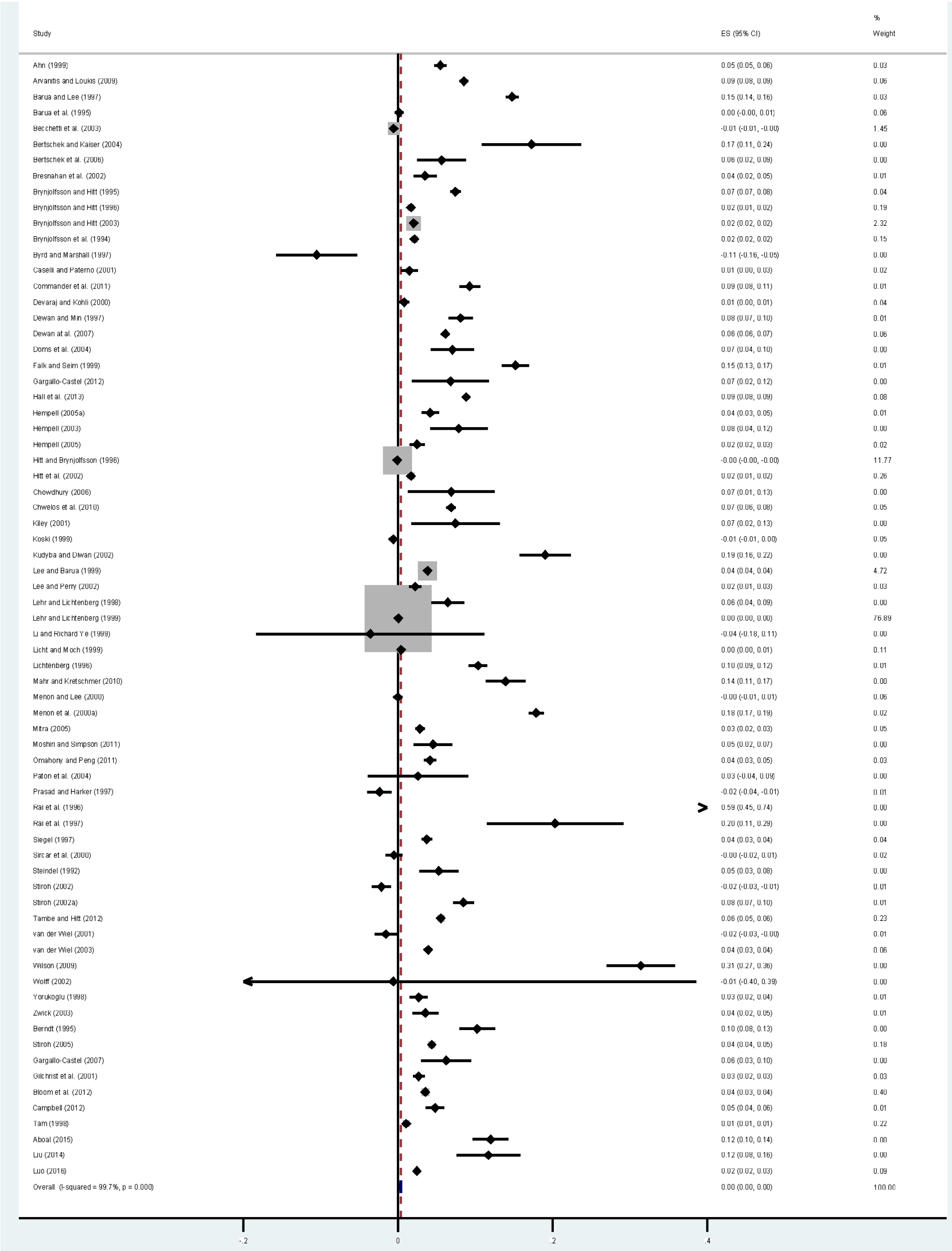


Figure 2.1: Forrest plot – weighted averages γ by study

2.6 Results

In case some of the estimates are more likely to be published, the simple arithmetic averages will deviate from the "true" value (Havranek, 2015). Figure 2.2 depicts the Epanechnikov kernel density of the estimates as a density plot, which is a commonly used analytical tool. The distribution of the estimates deviates from the normal distribution represented by the dashed line. Since the normal distribution is a standard assumption in the meta-analysis framework for the absence of publication bias, the striking difference between the plotted distribution of the estimates and the normal distribution leads to the suspicion of publication bias. The assumption of a normal distribution results from an econometric approach to the determination of the estimates by researchers. Due to the many sources of publication bias, its presence should be one of the assumption for every meta-analysis (Stanley, 2005, 2008). To reject or accept the hypothesis about the presence of publication bias, we will continue with graphical tests.

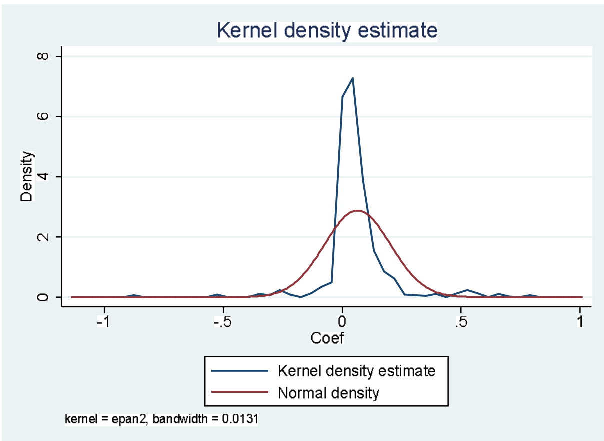
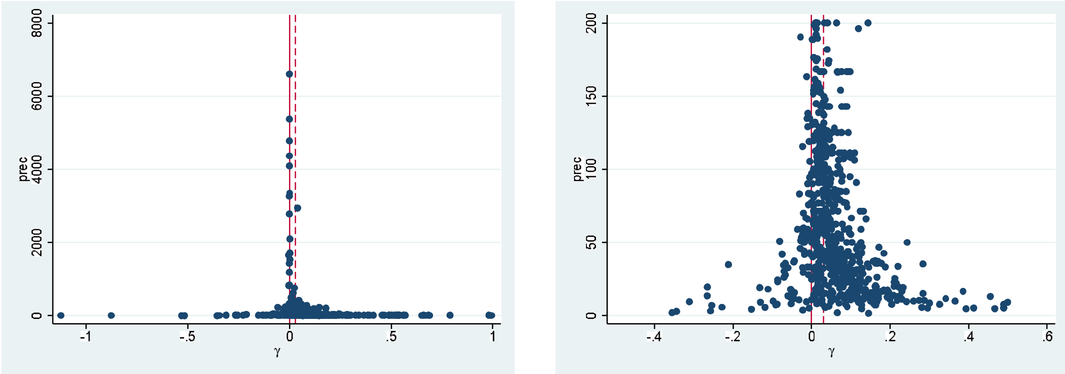


Figure 2.2: Kernel density of estimates

Graphical tests provide the easiest and fastest method for the detection of publication bias. First, funnel plots are depicted in Figure 2.3a for all studies and in detail in Figure 2.3b. The size of the estimates (on the x-axis) is plotted against its precision (inverted standard error) on the y-axis. The vertical dashed line depicts the value of 0.03, which represents the weighted average when using the inverse of the standard error for weighting the effect sizes. The detailed view allows us to observe that the estimates are more crowded on the positive part of the x-axis. A funnel plot usually identifies publication bias of type I – only estimates that fit the theory are accepted, as described in section 2.4.

Figure 2.3: Funnel plots of collected estimates

(a) Funnel graph – full sample (b) Funnel graph – detail

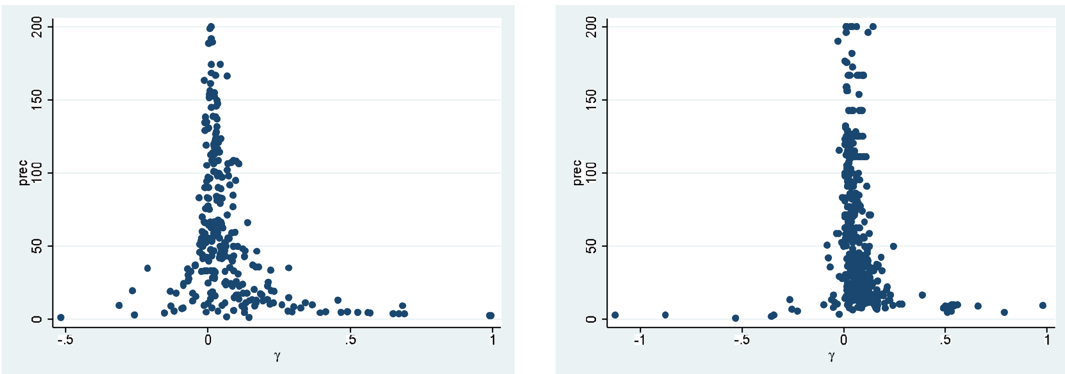


Notes: The funnel plot is a scatter plot where estimate (the ict elasticity) is plotted against its precision (inverse of standard error). The vertical solid line denotes the zero value, the dashed line denotes mean value.

To see the differences between the studies included in the previous meta-analysis and those included in this study, we can split the sample at year 2002 (as mentioned in chapter 2.5.1). These two groups are of about the same size, and the difference is clearly observable. On the one hand, the funnel graph in Figure 2.4a, which depicts estimates from studies written before 2002, is more or less symmetric. In contrast, Figure 2.4b depicts an obviously skewed funnel graph. The deviation from symmetry is mostly caused by "missing" negative effect sizes, which is consistent with type I publication bias.

Figure 2.4: Funnel plots of by publication year

(a) Funnel graph – before 2002 (b) Funnel graph – after 2002



Notes: The funnel plot is a scatter plot where estimate (the ict elasticity) is plotted against its precision (inverse of standard error). The year 2002 splits the sample into halfs.

Graphical testing of publication bias is clear, and possible selection can be easily revealed, but it is necessary to verify the visual results with precise econometric testing. Formal testing for publication bias use the same logic as graphical tests of funnel plots, as described in the previous section. We estimated model 2.7 and for robustness check, the OLS method with clustered standard errors was used (labelled as "Clustered OLS" in the tables). Table 2.3 summarizes the results. In all cases, the publication bias is positive and significant at the 1% significance level in all specifications. The magnitude of the publication bias is around 3, which according to Doucouliagos & Stanley (2008), means that the publication bias is so strong that it can produce significant results even if there is no true underlying effect. A positive and significant effect of ICT capital is found only for productivity using the mixed-effects method and reaching only 0.003, while the effect on profitability is not significantly different from zero. To estimate the true effect more precisely, we use model 2.9 proposed by Stanley & Doucouliagos (2007); the results are reported in table 2.4. We again use the mixed-effect method for estimation and the clustered OLS method with clustered standard errors for the robustness check.

Table 2.3: Publication bias testing

	Mixed-effects multilevel			Clustered OLS		
	Profitability	Productivity	All	Profitability	Productivity	All
prec	0.0000774	0.00262^{**}	0.00213^{**}	-0.00120^{*}	0.00235	0.00191
(effect size)	(0.09)	(5.10)	(5.20)	(-2.03)	(1.12)	(1.04)
Constant	2.977^{**}	3.228^{**}	3.141^{**}	3.165^{**}	2.747^{**}	2.831^{**}
(publication bias)	(6.15)	(4.44)	(5.81)	(12.68)	(10.18)	(12.42)
Observations	317	533	850	317	533	850

Dependent variable: tstat. t statistics in parentheses, for OLS clustered at the study level. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2.4: True ICT payoff

	Mixed-effects multilevel			Clustered OLS		
	Profitability	Productivity	All	Profitability	Productivity	All
prec	0.000911	0.00271^{**}	0.00221^{**}	0.00254	$0.00369^{\dagger}$	$0.00355^{\dagger}$
(true effect)	(1.00)	(5.28)	(5.40)	(1.29)	(1.69)	(1.86)
se	-9.701^{*}	-2.754	-3.837	15.43^{**}	7.275^{**}	8.874^{**}
	(-2.46)	(-0.79)	(-1.39)	(3.56)	(2.94)	(3.64)
Observations	317	533	850	317	533	850

Dependent variable: tstat. t statistics in parentheses, for OLS clustered at the study level. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

The results in table 2.4 are corrected for publication bias in line with the findings reported in table 2.3. The effect on profitability is found to be insignificantly different from zero, and the corrected effect on productivity is 0.003 and significant at the 1% level. This result is in strong contrast to the simple average of all estimates but is close to the weighted average of all studies.

We found publication bias in the ICT productivity literature with the highest reliable estimate of ICT elasticity being 0.003 and some estimates not being significant even at the 10% level. Our next research examines study characteristics as a source of the publication bias. The variables collected in the data gathering process will be also used in a simple explanatory meta-regression analysis (MRA), which aims at heterogeneity of the estimates. Disparateness between estimates from studies is given not only by some random error, but also by other factors including the data source or used methodology. Explaining the differences between studies was the main target of the previous meta-analyses Kohli & Devaraj (2003); Stiroh (2005), but neither of them considered publication bias and its sources. Compared to the previous studies, we use all estimates from every study, and even those estimates vary from model to model. One of the limitations of meta-analysis is that we have only a limited number of possible explanatory variables for this kind of differences, and we cannot capture all of it mainly because of the limited degrees of freedom. The variables used are selected with respect to the previous studies and summarized in Table 2.7. For heterogeneity modelling, we used the multilevel mixed effects model specified in model 2.8, which is best suited for explaining the diversity of the studies, for the publication bias we use model **??**.

The results of the multilevel mixed-effect model (model **??**) testing for publication bias sources are reported in table 2.5. The results of the explanatory meta-regression analysis (model 2.8) are reported in table 2.8, both showing the effects on productivity and profitability. One might think that the interpretation of the results is not straightforward because the dependent variable is the t-statistic. Becasue the coefficients of the explanatory variables would not provide the magnitude of the effect and one can only interpret the sign and significance. In case the coefficient is negative, the corresponding variable underestimates the effect, and if the coefficient is positive, it results in an overestimation.

The findings do not reveal much direct information about the factors that influence the publication bias: only a few of the gathered explanatory variables turned out to be significant and influence the research outcomes. The number

Table 2.5: Meta-regression analysis: Sources of publication bias

	Mixed-effects multilevel			Clustered OLS		
	Profitability	Productivity	All	Profitability	Productivity	All
prec	-0.000150	0.00308**	0.00229**	0.0000229	0.00236	0.00173
	(-0.18)	(5.88)	(5.56)	(0.02)	(0.81)	(0.71)
Observations	1.408**	0.422	0.795**	0.927**	0.0780	0.431
	(7.57)	(1.45)	(3.75)	(3.69)	(0.16)	(1.57)
Years	0.155**	-0.247*	0.0190	0.103†	-0.213	-0.0353
	(3.14)	(-2.14)	(0.27)	(1.82)	(-1.14)	(-0.31)
Labour	-5.091*	-1.233	-2.106*	-6.599**	-0.0977	-0.686
	(-2.43)	(-0.99)	(-2.16)	(-2.89)	(-0.11)	(-0.82)
Publication	-0.542	0.359	-0.136	-0.494*	0.226	-0.0875
year	(-1.40)	(0.67)	(-0.30)	(-2.34)	(0.56)	(-0.31)
Publication	0.0125	0.0115	0.0113	0.0123	0.0178	0.0120
year^2	(0.88)	(0.60)	(0.71)	(1.52)	(1.54)	(1.59)
Citations	-0.0154	0.310	0.184	0.167	0.455*	0.297†
	(-0.05)	(0.61)	(0.44)	(0.69)	(2.04)	(1.80)
Working paper	-0.515	0.760	-0.0523	-0.0400	1.542	0.802
	(-0.52)	(0.34)	(-0.03)	(-0.05)	(1.06)	(0.95)
OLS method	0.383	-0.0887	0.0286	1.593*	0.126	0.752
	(0.94)	(-0.12)	(0.06)	(2.46)	(0.15)	(1.12)
countryUS	2.300*	0.984	0.590	1.868*	1.330	0.637
	(2.31)	(0.62)	(0.50)	(2.69)	(1.15)	(0.99)
Average year	0.183*	-0.506**	-0.135	0.147†	-0.439	-0.118
	(2.53)	(-3.57)	(-1.50)	(1.91)	(-1.33)	(-0.62)
Impact factor	-0.468	-1.468	-0.625	-0.663*	-1.320†	-0.383
	(-0.96)	(-1.51)	(-0.90)	(-2.28)	(-1.69)	(-1.29)
Dependent			-0.780			-0.374
variable			(-1.23)			(-0.57)
Constant	-369.5*	1004.6**	268.0	-293.9†	873.2	234.7
	(-2.56)	(3.57)	(1.50)	(-1.92)	(1.33)	(0.62)
Observations	317	533	850	317	533	850
rmse				2.500	6.242	5.290

t statistics in parentheses, for OLS clustered at the study level

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

Dependent variable: tstat

of observations and data source used in the regression for estimation is found to be meaningful. The higher the number of observations (*nobs*), the higher the publication bias. The same holds for US data when examining profitability, but there is not a difference in productivity. At the 10% significance level, the specification for profitability finds lower bias when the dependent variable is normalized with labour. Interestingly, the length of the period used for estimating the ICT effects as well as the average year (more recent studies) decrease the bias for productivity but increase for profitability estimates.

We found no difference between working papers and published works. The number of citation references, the methodology used and the impact factor of the journal are also not related to the publication bias. This leads to an interesting conclusion about the source of the publication bias – the difference between working papers and published papers is insignificant by any tested variable, therefore I can claim that publication bias in ICT literature is not caused by pressure from editors or referees to publish significant results (in other words, working papers and journal articles do not differ in the extent of publication bias). Instead, the bias seems to stem from self-censorship by authors, who often view negative and insignificant results as either wrong or uninteresting.

2.6.1 Determinants of ICT payoff

Previous meta-analyses were aimed at explaining heterogeneity between studies, and for this reason, the results of the performed MRA are comparable to the previous findings in the literature. Table 2.6 captures the main findings of two previous meta-analyses and compares them to results of this paper (see table 2.8 for details). As mentioned, the focus of the previous studies was a bit different, and thus, we did not included all of their propositions.

First, difference was found between OLS and the other methods of estimation only for profitability estimates. Positive relationship between impact factor and the effect size is identified for productivity, authors with lower estimates target journals of higher quality. But there are no differences between the estimates from the working papers and those from the published studies. This indicates that the effect size is not related to the "popularity" of the estimate. The hypothesis that the data from the US will lead to a higher effect size is partially supported in the results. A positive effect is found only for the profitability estimates.

Table 2.6: Summary and comparison of findings

Proposition	Kohli (2003)	Stiroh (2005)	This paper (2017)
IT payoffs differ among industry sectors	Supported	Mixed results	–
Larger sample size leads to greater ICT payoff	Supported	–	Mixed results
Estimates of the profitability-based dependent variable differ from the productivity-based estimates	Supported	Not supported	Strongly Supported
Labour productivity differs from not-per-labour normalized dependent variable	–	Not supported	Mixed results*
Longitudinal estimates are higher than short-term estimates	Not supported	–	Mixed results*
Estimates based on more recent data show higher estimates	–	Supported	Mixed results
There is publication bias in the ICT payoff literature	–	–	Supported
Genuine effect size	–	0.06	0.002 – 0.003

* proposition not supported by all specifications

Kohli & Devaraj (2003) concludes that a larger sample size leads to a higher ICT payoff, but we find different effects for profitability and productivity estimates. Next to it, we agree with the conclusion that "studies with profitability based dependent variable have different ICT payoff than those that measure productivity" (Kohli & Devaraj, 2003, 130). With Stiroh (2005), we disagree that no significant differences appear between the estimates of productivity and the profitability. He also finds a positive relation between the average year in the dataset and the effect size of the ICT investment, but our analysis finds a negative relationship when the data were analysed using a multilevel model for productivity estimates. The rest of the results are mixed – we do not have clearly contradicting or supporting results.

What implications can be drawn from this paper? First, the effects of ICT investment on productivity are lower than commonly expected. Second, the productivity paradox literature carries a burden called publication bias that causes overestimation of the ICT elasticity. Some of the results also show the possibility that the productivity paradox is "reborn". It seems that the method of estimation does not affect the results for productivity estimates, and therefore, either the general Cobb-Douglas production function framework is not proper for investigating ICT effects or the model should incorporate possible explanations of the productivity paradox as a control variable. If ICT does not increase productivity per se, then any company should make a proper case study before investing into ICT because otherwise it will probably result in misinvesting.

2.6.2 Limitations and Future Research

The results presented in this paper do have some limitations that stem from the used methods. Some of them can be challenged in future research. First and foremost, as with every meta-analysis, the main limitations are the data availability and quality. If the underlying studies are properly carried out, then the conclusions of the meta-analysis are also more reliable. As my sample contains more than 70 studies and more than two thirds of the used studies were published, the estimates should reach sufficient quality. It is enough for publication bias testing, but for explanatory MRA, the characteristics of the studies and the estimates are needed to capture the differences in quality.

Strong publication bias was found, but most of the variables used to explain it were found to be insignificant. Further research should therefore try to identify the key drivers of this bias. Overall, MRA was able to explain only a small portion of the variability of the results. Thus, finding additional explanatory variables that would help to increase the proportion of explained variability is needed. This could also eliminate the mixed results we obtained. We employed several models to check for the robustness of the estimated numbers. The use of a wide variety of methods usually helps to check for robustness of the results and to reduce doubts about the reliability of the estimates, but in the case of this research, we found mixed results about the signs, which lead to ambiguous conclusions.

It is also questionable whether we can draw some really general and broad recommendations for every company. Most of the studies are based on US data (63% US, 29% Europe, 8% other). Therefore, the implications for Europe might be different. This study also investigated only firm level and direct effects of ICT. Aggregated levels should also be investigated to determine the significance of public spending and test for spillover effects. We also should not forget that ICT capital requires skilled workers. A proposition for future research is to perform a meta-analysis of the relation between investments into ICT workers and firm performance. Positive results would not only be intuitive but also consistent with theory and have relevance for business decisions.

2.7 Conclusion

The central topic of this paper is to assess the productivity paradox from a meta-analytical perspective with a focus on publication bias and with the aim of revealing the genuine effect of ICT investments on productivity at the firm level, which makes this paper unique. Meta-analysis is a quite powerful and widely used tool for the synthesis of empirical research findings, but specific methods differ. This paper is designed accordingly. The first part reviews the ICT productivity paradox and the main literature findings. The second part is devoted to meta-analysis and the techniques used for publication bias investigation, followed by the third empirical part, where the data gathered across the available literature are analysed. Emphasis is put on multilevel analysis because we use all of the estimates from the studies and the estimation of the genuine effect after accounting for the publication bias. The last part of the paper concludes.

The productivity paradox has been investigated for decades, and from the large amount of published literature, one can hardly expect to make a general conclusion without using proper summarizing techniques. We employ meta-analysis to find the true effect of ICT investments on productivity. For that purpose, more than 850 estimates from more than 70 studies were collected together with some other descriptive indicators. The existing literature has focused on finding a positive effect to refute the paradox. Later, the diversity of the results was addressed by Kohli & Devaraj's 2003 meta-analysis, and Stiroh (2005) found the ICT payoff to be around 0.06 when using a mixture of firm and aggregated level literature and one estimate per study.

This paper finds clear and substantial evidence of publication bias presence in the ICT productivity paradox literature. By filtering this bias, the underlying effect is identified to be around 0.003 when the data are analysed using a mixed-effects model. If we combine the last finding with evidence from MRA showing that ICT elasticity decreases with increasing average year in the dataset, it appears that the productivity paradox might be reborn after it was refuted. A possible explanation could be the fact that in today's PC driven world, technology is a must, and thus, ICT technology is so incorporated into any capital and production technology that we cannot clearly separate it and find positive effects of ICT-related investments. Another explanation could be the investment into ICT, which lowers the resulting effect. The main drivers of inadequate investment are mostly wrong management, but it can also be the impact of fashions in ICT.[5] Managers are evaluated better when chasing the newest ICT, and thus, they have high incentive to invest in such technologies even if the payoff is insignificant.

Strong publication bias is identified, but the explanatory variables related to publication bias are found to be mostly insignificant. We find no evidence of different results between working papers and published studies, and no effect of journal quality or citations concluding that publication bias is not caused by the review and publication process, but by the self-censorship of researches. Several studies have suggested including the square of publication year to control for the so-called research cycle hypothesis connected with publication bias, but no support for the this hypothesis has been found. The limitations of this paper are based on the ability to explain the diversity of results from only a few and quite general descriptive variables about each study whose estimates can be retrieved from the studies and coded. Since the size of the effect helps make the right decision in business-related investments, our result of ICT elasticity being very close to zero with values, about 0.3% for productivity and no effect on profitability, supports the argument that there are better forms of investment to be made.

[5] "An IT fashion is a transitory collective belief that an information technology is new, efficient, and at the forefront of practice." (Wang, 2010, 63)

Bibliography

Aboal, D. & E. Tacsir (2015): "Innovation and productivity in services and manufacturing: The role of ict investment." *Technical report*, Inter-American Development Bank.

Ahn, S. (1999): "Technology upgrading with learning cost: A solution for two 'productivity puzzles'." *Technical Report 220*, OECD Economics Department Working Papers.

Alshilash, A. A. (1997): *A study to measure the use and effects of information technology on organizational functions*. Ph.D. thesis, George Washington University, Washington, D.C.

Arvanitis, S. & E. N. Loukis (2009): "Information and communication technologies, human capital, workplace organization and labour productivity: A comparative study based on firm-level data for greece and switzerland." *Information Economics and Policy* **21(1)**: pp. 43–61.

Ashenfelter, O. & M. Greenstone (2004): "Estimating the value of a statistical life: The importance of omitted variables and publication bias." *The American Economic Review* **94(2)**: pp. 454–460.

Atrostic, B. K. & S. V. Nguyen (2005): "It and productivity in u.s. manufacturing: Do computer networks matter?" *Economic Inquiry* **43(3)**: pp. 493–506.

Baily, M. N., R. J. Gordon, W. D. Nordhaus, & D. Romer (1988): "The productivity slowdown, measurement issues, and the explosion of computer power." *Brookings Papers on Economic Activity* **1988(2)**: pp. 347–431.

Barua, A., C. H. Kriebel, & T. Mukhopadhyay (1995): "Information technologies and business value: An analytic and empirical investigation." *Information Systems Research* **6(1)**: pp. 3–23.

Barua, A. & B. Lee (1997): "The information technology productivity paradox revisited: A theoretical and empirical investigation in the manufacturing sector." *International Journal of Flexible Manufacturing Systems* **9(2)**: pp. 145–166.

Bateman, I. J. & A. P. Jones (2003): "Contrasting conventional with multilevel modeling approaches to meta-analysis: Expectation consistency in u.k. woodland recreation values." *Land Economics* **79(2)**: pp. 235–258.

Becchetti, L., D. Bedoya, & L. Paganetto (2003): "Ict investment, productivity and efficiency: Evidence at firm level using a stochastic frontier approach." *Journal of Productivity Analysis* **20(2)**: pp. 143–167.

Berndt, E. R. & C. J. Morrison (1995): "High-tech capital formation and economic performance in u.s. manufacturing industries an exploratory analysis." *Journal of Econometrics* **65(1)**: pp. 9–43.

Bertschek, I. & U. Kaiser (2004): "Productivity effects of organizational change: Microeconometric evidence." *Management Science* **50(3)**: pp. 394–404.

Bertschek, I., U. Kaiser, & H. Fryges (2006): "B2b or not to be: Does b2b e-commerce increase labour productivity?" *International Journal of the Economics of Business* **13(3)**: pp. 387–405.

Biagi, F. *et al.* (2013): "Ict and productivity: A review of the literature." *Institute for Prospective Technological Studies. Digital Economy Working Paper* **9**.

Bickel, R. (2007): *Multilevel Analysis for Applied Research: It's Just Regression!* Guilford Publications.

Bloom, N., R. Sadun, & J. Van Reenen (2012): "Americans do it better: Us multinationals and the productivity miracle." *The American Economic Review* **102(1)**: pp. 167–201.

Bresnahan, T. F., E. Brynjolfsson, & L. M. Hitt (2002): "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence." *The Quarterly Journal of Economics* **117(1)**: pp. 339–376.

Brynjolfsson, E. (1993): "The productivity paradox of information technology." *Communications of the ACM* **36(12)**: pp. 66–77.

Brynjolfsson, E. & L. M. Hitt (1994): "Computers and economic growth : firm-level evidence." *Technical Report 3714-94. CISR WP ; no. 27*, MIT.

Brynjolfsson, E. & L. M. Hitt (1995): "Information technology as a factor of production: The role of differences among firms." *Economics of Innovation and New Technology* **3(3-4)**: pp. 183–200.

Brynjolfsson, E. & L. M. Hitt (1996): "Paradox lost? firm-level evidence on the returns to information systems spending." *Management Science* **42(4)**: pp. 541–558.

Brynjolfsson, E. & L. M. Hitt (1998): "Beyond the productivity paradox." *Communications of the ACM* **41(8)**: pp. 49–55.

Brynjolfsson, E. & L. M. Hitt (2000): "Beyond computation: Information technology, organizational transformation and business performance." *The Journal of Economic Perspectives* **14(4)**: pp. 23–48.

Brynjolfsson, E. & L. M. Hitt (2003): "Computing productivity: Firm-level evidence." *The Review of Economics and Statistics* **85(4)**: pp. 793–808.

Byrd, T. & T. Marshall (1997): "Relating information technology investment to organizational performance: a causal model analysis." *Omega* **25(1)**: pp. 43–56.

Campbell, M. (2012): "What a difference a year makes: Time lag effect of information technology investment on firm performance." *Journal of Organizational Computing and Electronic Commerce* **22(3)**: pp. 237–255.

Card, D. & A. B. Krueger (1995): "Time-series minimum-wage studies: A meta-analysis." *The American Economic Review* **85(2)**: pp. 238–243.

Cardona, M., T. Kretschmer, & T. Strobel (2013): "Ict and productivity: conclusions from the empirical literature." *Information Economics and Policy* **25(3)**: pp. 109–125.

Caselli, P. & F. Paterno (2001): "Ict accumulation and productivity growth in the united states: an analysis based on industry data." *Temi di discussione (Economic working papers) 419*, Bank of Italy, Economic Research and International Relations Area.

Chen, W.-Y. (1997): *On the relationshipbetween information technology investment and corporate performance: An empirical investigation of the United States industries*. Ph.D. thesis, Drexel University, Philadelphia. Unpublished Ph.D. dissertation.

Chowdhury, S. K. (2006): "Investments in ict-capital and economic performance of small and medium scale enterprises in east africa." *Journal of International Development* **18(4)**: pp. 533–552.

Christensen, P. (2003): "Topics in meta-analysis. a literature survey." *Technical report*, Institute of Transport Economics, Oslo.

Chwelos, P., R. Ramirez, K. L. Kraemer, & N. P. Melville (2010): "Research note–does technological progress alter the nature of information technology as a production input? new evidence and new results." *Information Systems Research* **21(2)**: pp. 392–408.

Cline, M. K. (1999): *The impact of information technology investment on productivity improvements in the American railroad industry between 1986 and 1995*. Ph.D. thesis, Florida State University.

Commander, S., R. Harrison, & N. Menezes-Filho (2011): "Ict and productivity in developing countries: New firm-level evidence from brazil and india." *The Review of Economics and Statistics* **93(2)**: pp. 528–541.

David, P. A. (1990): "The dynamo and the computer: An historical perspective on the modern productivity paradox." *American Economic Review* **80(2)**: pp. 355–61.

David, P. A. (2000): *Understanding the digital economy: Data, tools, and research*, chapter Understanding digital technology's evolution and the path of measured productivity growth: Present and future in the mirror of the past, pp. 49–95. Mit Press.

Dedrick, J., V. Gurbaxani, & K. Kraemer (2003): "Information technology and economic performance: A critical review of the empirical evidence." *ACM Computing Surveys (CSUR)* **35(1)**: pp. 1–28.

Devaraj, S. & R. Kohli (2000): "Information technology payoff in the healthcare industry: a longitudinal study." *Journal of Management Information Systems* **16(4)**: pp. 41–67.

Devaraj, S. & R. Kohli (2003): "Performance impacts of information technology: Is actual usage the missing link?" *Management science* **49(3)**: pp. 273–289.

DEWAN, S. & C.-k. MIN (1997): "The substitution of information technology for other factors of production: A firm level analysis." *Management Science* **43(12)**: pp. 1660–1675.

DEWAN, S., C. SHI, & V. GURBAXANI (2007): "Investigating the risk-return relationship of information technology investment: Firm-level empirical analysis." *Management Science* **53(12)**: pp. 1829–1842.

DOMS, M. E., R. S. JARMIN, & S. D. KLIMEK (2004): "Information technology investment and firm performance in us retail trade." *Economics of Innovation and New Technology* **13(7)**: pp. 595–613.

DOUCOULIAGOS, C. & P. LAROCHE (2009): "Unions and profits: A meta-regression analysis." *Industrial Relations: A Journal of Economy and Society* **48(1)**: pp. 146–184.

DOUCOULIAGOS, C. & T. D. STANLEY (2008): "Theory competition and selectivity: Are all economic facts greatly exaggerated?" *Technical Report 6.*

DOUCOULIAGOS, C. & T. D. STANLEY (2011): "Are all economic facts greatly exaggerated? theory competition and selectivity." *Journal of Economic Surveys* **0**: pp. 1–29.

DRACA, M., R. SADUN, & J. V. REENEN (2006): "Productivity and ict: A review of the evidence." *Technical Report dp0749*, CEP Discussion Papers.

EGGER, M., D. G. SMITH, M. SCHNEIDER, & C. MINDER (1997): "Bias in meta-analysis detected by a simple, graphical test." *British Medical Journal* **315(7109)**: pp. 629–634.

FALK, M. & K. SEIM (1999): "Workers' skill level and information technology: evidence from german service firms." *Technical Report 99-14*, ZEW Discussion Papers.

GARGALLO-CASTEL, A. F. & C. GALVE-GÓRRIZ (2007): "Information technology, complementarities and three measures of organizational performance: Empirical evidence from spain." *Journal of Information Technology Impact* **7(1)**: pp. 43–58.

GARGALLO-CASTEL, A. F. & C. GALVE-GÓRRIZ (2012): "A firm-level investigation of the complementarity between information and communication technologies and organizational resources." *Procedia - Social and Behavioral Sciences* **41(0)**: pp. 51–58.

Gelman, A. & J. Hill (2006): *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.

Gilchrist, S., V. Gurbaxani, & R. Town (2001): "Productivity and the pc revolution." *Center for Research on Information Technology and Organizations Working paper, University of California* **12**: pp. –.

Glass, G. V. (1976): "Primary, secondary, and meta-analysis of research." *Educational Researcher* **5(10)**: pp. 3–8.

Grover, V., J. Teng, A. H. Segars, & K. Fiedler (1998): "The influence of information technology diffusion and business process change on perceived productivity: The is executive's perspective." *Information & Management* **34(3)**: pp. 141–159.

Hall, B. H., F. Lotti, & J. Mairesse (2013): "Evidence on the impact of r&d and ict investments on innovation and productivity in italian firms." *Economics of Innovation and New Technology* **22(3)**: pp. 300–328.

Han, K., Y. Chang, & J. Hahn (2011): "Information technology spillover and productivity: The role of information technology intensity and competition." *J. Manage. Inf. Syst.* **28(1)**: pp. 115–146.

Harris, D. (1994): *Organizational linkages: Understanding the productivity paradox*. National Academies Press.

Havranek, T. (2010): "Rose effect and the euro: Is the magic gone?" *Review of World Economics* **146(2)**: pp. 241–261.

Havranek, T. (2015): "Measuring intertemporal substitution: The importance of method choices and selective reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

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 (2016): "Do borders really slash trade? a meta-analysis." **(forthcoming)**.

Hempell, T. (2003): "Do computers call for training? firm-level evidence on complementarities between ict and human capital investments." *Technical Report 03-20*, ZEW Discussion Papers.

Hempell, T. (2005a): "Does experience matter? innovations and the productivity of information and communication technologies in german services." *Economics of Innovation and New Technology* **14(4)**: pp. 277–303.

Hempell, T. (2005b): "What's spurious, what's real? measuring the productivity impacts of ict at the firm-level." *Empirical Economics* **30(2)**: pp. 427–464.

Hitt, L. M. & E. Brynjolfsson (1996): "Productivity, business profitability, and consumer surplus: Three different measures of information technology value." *MIS Quarterly* **20(2)**: pp. 121–142.

Hitt, L. M., D. J. Wu, & X. Zhou (2002): "Investment in enterprise resource planning: Business impact and productivity measures." *Journal of Management Information Systems* **19(1)**: pp. 71–98.

Horzella, s. (2005): *Beyond IT and productivity - Effects of Digitized Information Flows in Grocery Distribution*. Ph.D. thesis, Linköpings universitet.

Hox, J. (1995): *Applied multilevel analysis*. TT-publikaties, Amsterdam.

Hox, J. & E. Leeuw (2003): "Multilevel models for meta-analysis." In "Reise SP, Duan N (eds) Multilevel modeling: methodological advances and applications," pp. 90–111. Lawrence Erlbaum Associates Publishers.

Hunter, J. & F. Schmidt (2004): *Methods of meta-analysis: correcting error and bias in research findings*. Sage, 2 edition.

Johnston, R., E. Besedin, & R. Wardwell (2003): "Modeling relationships between use and nonuse values for surface water quality: A meta-analysis." *Water Resources Research* **39(12)**: pp. 1363–.

Kelley, M. R. (1994): "Productivity and information technology: The elusive connection." *Management Science* **40(11)**: pp. 1406–1425.

KILEY, M. T. (2001): "Computers and growth with frictions: aggregate and disaggregate evidence." *Carnegie-Rochester Conference Series on Public Policy* **55(1)**: pp. 171–215.

KIVIJÄRVI, H. & T. SAARINEN (1995): "Investment in information systems and the financial performance of the firm." *Information & Management* **28(2)**: pp. 143–163.

KOHLI, R. & S. DEVARAJ (2003): "Measuring information technology payoff: A meta-analysis of structural variables in firm-level empirical research." *Info. Sys. Research* **14(2)**: pp. 127–145.

KOSKI, H. (1999): "The implications of network use, production network externalities and public networking programmes for firm's productivity." *Research Policy* **28(4)**: pp. 423–439.

KUDYBA, S. & R. DIWAN (2002): "The impact of information technology on us industry." *Japan and the World Economy* **14(3)**: pp. 321–333.

LEE, B. & A. BARUA (1999): "An integrated assessment of productivity and efficiency impacts of information technology investments: Old data, new analysis and evidence." *Journal of Productivity Analysis* **12(1)**: pp. 21–43.

LEE, G. & J. L. PERRY (2002): "Are computers boosting productivity? a test of the paradox in state governments." *Journal of Public Administration Research and Theory* **12(1)**: pp. 77–102.

LEHR, B. & F. R. LICHTENBERG (1999): "Information technology and its impact on firm-level productivity: evidence from government and private data sources, 1977-1993." *Canadian Journal of Economics* **32(2)**: pp. 335–362.

LEHR, W. & F. R. LICHTENBERG (1998): "Computer use and productivity growth in us federal government agencies, 1987-92." *The Journal of Industrial Economics* **46(2)**: pp. 257–279.

LI, M. & L. RICHARD YE (1999): "Information technology and firm performance: Linking with environmental, strategic and managerial contexts." *Information & Management* **35(1)**: pp. 43–51.

LICHT, G. & D. MOCH (1999): "Innovation and information technology in services." *The Canadian Journal of Economics* **32(2)**: pp. 363–383.

Lichtenberg, F. R. (1996): "The output contributions of computer equipment and personnel: A firm- level analysis." *Technical report*, NBER Working Paper Series.

Lim, J.-H., B. Dehning, V. J. Richardson, & R. E. Smith (2011): "A meta-analysis of the effects of it investment on firm financial performance." *Journal of Information Systems* **25(2)**: pp. 145–169.

Lim, J.-H., V. J. Richardson, & T. L. Roberts (2004): "Information technology investment and firm performance: A meta-analysis." In "Proceedings of the 37th Hawaii International Conference on System Sciences," volume 8, pp. 80221a–80221a.

Liu, T.-K., J.-R. Chen, C. J. Huang, & C.-H. Yang (2014): "Revisiting the productivity paradox: A semiparametric smooth coefficient approach based on evidence from taiwan." *Technological Forecasting and Social Change* **81**: pp. 300–308.

Long, J. B. D. & K. Lang (1992): "Are all economic hypotheses false?" *Journal of Political Economy* **100(6)**: pp. 1257–1272.

Lubbe, S., G. Parker, & A. Hoard (1995): "The profit impact of it investment." *Journal of Information Technology (Routledge, Ltd.)* **10(1)**: pp. 44–.

Luo, Y. & J. Bu (2016): "How valuable is information and communication technology? a study of emerging economy enterprises." *Journal of World Business* **51(2)**: pp. 200 – 211.

Mahmood, M. A. & G. J. Mann (1993): "Impact of information technology investment: An empirical assessment." *Accounting, Management and Information Technologies* **3(1)**: pp. 23–32.

Mahr, F. & T. Kretschmer (2010): "Complementarities between it and organizational structure: The role of corporate exploration and exploitation." *Technical Report 11507*, Discussion Papers in Business Administration.

Mayberry-Stewart, M. I. T. (1996): *Determinants of information technology strategic performance: An analysis of firm performance, IT performance, IT investment intensity and IT governance in the health care sector*. Ph.D. thesis, Claremont Graduate School, Claremont.

Menon, N. M. & B. Lee (2000): "Cost control and production performance enhancement by it investment and regulation changes: evidence from the healthcare industry." *Decision Support Systems* **30(2)**: pp. 153–169.

Menon, N. M., B. Lee, & L. Eldenburg (2000): "Productivity of information systems in the healthcare industry." *Information Systems Research* **11(1)**: pp. 83–92.

Mitra, S. (2005): "Information technology as an enabler of growth in firms: An empirical assessment." *Journal of Management Information Systems* **22(2)**: pp. 279–300.

Mittal, N. & B. Nault (2009): "Research note—investments in information technology: Indirect effects and information technology intensity." *Information Systems Research* **20(1)**: pp. 140–154.

Moshiri, S. & W. Simpson (2011): "Information technology and the changing workplace in canada: firm-level evidence." *Industrial and Corporate Change* **20(6)**: pp. 1601–1636.

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(3)**: pp. 345–377.

Oliner, S. D. & D. E. Sichel (2000): "The resurgence of growth in the late 1990s: Is information technology the story?" *Journal of Economic Perspectives* **14(4)**: pp. 3–22.

O'Mahony, M. & F. Peng (2011): "Intangible training capital and productivity in europe." *Technical Report 38648*, MPRA Paper.

Paton, D., D. S. Siegel, & L. V. Williams (2004): "Productivity measurement in a service industry: Plant-level evidence from gambling establishments in the united kingdom." *Technical Report 0413*, Rensselaer Working Papers in Economics.

Peters, J. L., A. J. Sutton, D. R. Jones, K. R. Abrams, L. Rushton, & S. G. Moreno (2010): "Assessing publication bias in meta-analyses in the presence of between-study heterogeneity." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **173(3)**: pp. 575–591.

PRASAD, B. & P. HARKER (1997): "Examining the contribution of information technology toward productivity and profitability in us retail banking." *The Wharton Financial Institutions Center Working Papers* **97(9)**: pp. –.

RAI, A., R. PATNAYAKUNI, & N. PATNAYAKUNI (1996): "Refocusing where and how it value is realized: An empirical investigation." *Omega* **24(4)**: pp. 399–412.

RAI, A., R. PATNAYAKUNI, & N. PATNAYAKUNI (1997): "Technology investment and business performance." *Communications of the ACM* **40(7)**: pp. 89–97.

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. & 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.

SIEGEL, D. (1997): "The impact of computers on manufacturing productivity growth: A multiple-indicators, multiple-causes approach." *Review of Economics and Statistics* **79(1)**: pp. 68–78.

SIRCAR, S., J. L. TURNBOW, & B. BORDOLOI (2000): "A framework for assessing the relationship between information technology investments and firm performance." *Journal of Management Information Systems* **16(4)**: pp. 69–97.

SOLOW, R. M. (1957): "Technical change and the aggregate production function." *The Review of Economics and Statistics* **39(3)**: pp. 312–320.

SOLOW, R. M. (1987): "We'd better watch out." In "New York Times Book Review," New York Times Book Review, p. 36.

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. & C. DOUCOULIAGOS (2007): "Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression." *Economics Series* **11**: p. 2007.

STANLEY, T. D. & C. 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. & C. DOUCOULIAGOS (2014): "Meta-regression approximations to reduce publication selection bias." *Research Synthesis Methods* **5(1)**: pp. 60–78.

STANLEY, T. D., C. 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.

STEINDEL, C. (1992): "Manufacturing productivity and high-tech investment." *FRBNY Quarterly Review* **Summer**: pp. 39–47.

STERNE, J. A. C., M. EGGER, & G. D. SMITH (2001): "Investigating and dealing with publication and other biases in meta-analysis." *BMJ* **323**: pp. –.

STERNE, J. A. C., D. GAVAGHAN, & M. EGGER (2000): "Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature." *Journal of Clinical Epidemiology* **53(11)**: pp. 1119–1129.

STIROH, K. J. (2002a): "Are ict spillovers driving the new economy?" *Review of Income and Wealth* **48(1)**: pp. 33–57.

STIROH, K. J. (2002b): "Information technology and the u.s. productivity revival: What do the industry data say?" *The American Economic Review* **92(5)**: pp. 1559–1576.

STIROH, K. J. (2005): "Reassessing the impact of it in the production function: A meta-analysis and sensitivity tests." *Annales d'Economie et de Statistique* **(79-80)**: pp. 529–561.

TAM, K. Y. (1998): "Analysis of firm-level computer investments: A comparative study of three pacific-rim economies." *IEEE Transactions on Engineering Management* **45(3)**: pp. 276–286.

TAMBE, P. & L. M. HITT (2012): "The productivity of information technology investments: New evidence from it labor data." *Information Systems Research* **23(3-Part-1)**: pp. 599–617.

TEO, T. S., P. K. WONG, & E. HUI CHIA (2000): "Information technology (it) investment and the role of a firm: an exploratory study." *International Journal of Information Management* **20(4)**: pp. 269–286.

THATCHER, M. & J. OLIVER (2001): "The impact of technology investments on a firm's production efficiency, product quality, and productivity." *Journal of Management Information Systems* **18(2)**: pp. 17–46.

TRIPLETT, J. E. (1999): "The solow productivity paradox: what do computers do to productivity?" *Canadian Journal of Economics* **32(2)**: pp. 309–334.

TURBAN, E., E. MCLEAN, & J. WETHERBE (2002): *Information technology for management: transforming business in the digital economy.* J. Wiley, 3 edition.

VENTURINI, F. (2009): "The long-run impact of ict." *Empirical Economics* **37(3)**: pp. 497–515.

WANG, P. (2010): "Chasing the hottest it: effects of information technology fashion on organizations." *MIS quarterly* **34(1)**: pp. 63–85.

VAN DER WIEL, H. (2001): "Does ict boost dutch productivity growth?" *Technical Report 16*, CPB Netherlands Bureau for Economic Policy Analysis.

VAN DER WIEL, H. & G. VAN LEEUWEN (2003): "Do ict spillovers matter; evidence from dutch firm-level data." *Technical Report 26*, CPB Netherlands Bureau for Economic Policy Analysis.

WILLCOCKS, L. & S. LESTER (1996): "Beyond the it productivity paradox." *European Management Journal* **14(3)**: pp. 279–290.

WILSON, D. J. (2009): "It and beyond: The contribution of heterogeneous capital to productivity." *Journal of Business & Economic Statistics* **27(1)**: pp. 52–70.

WOLFF, E. N. (2002): "Productivity, computerization, and skill change." *National Bureau of Economic Research Working Paper Series* **No. 8743**: pp. –.

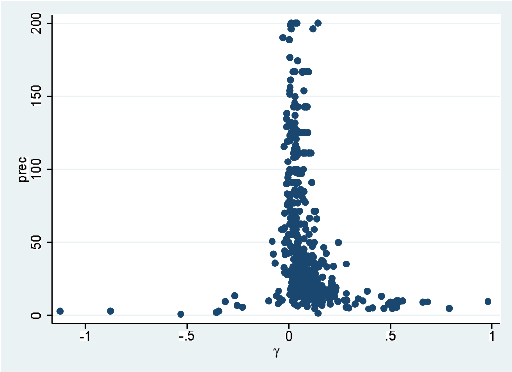
YORUKOGLU, M. (1998): "The information technology productivity paradox." *Review of Economic Dynamics* **1(2)**: pp. 551–592.

ZWICK, T. (2003): "The impact of ict investment on establishment productivity." *National Institute Economic Review* **184(1)**: pp. 99–110.

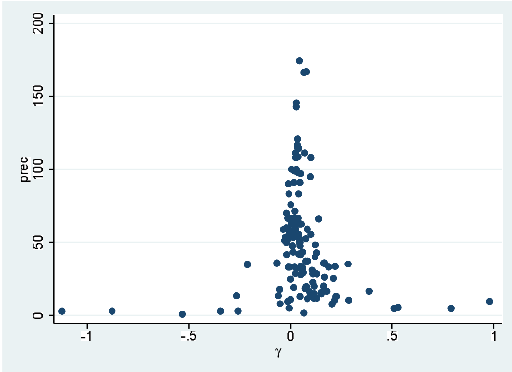
2.A Appendix

Figure 2.5: Funnel plots for new studies and working papers

(a) Funnel graph – new studies



(b) Funnel graph – working papers



Notes: The funnel plot is a scatter plot where estimate (the ict elasticity) is plotted against its precision (inverse of standard error).

Table 2.7: Explanatory variables used in regressions

Variable	Description
gamma	point estimate of effect of ICT on productivity
se	standard error of the estimate γ
prec	inverse of *se*
Observations	logarithm of number of observations the estimate is calculated from
tstat	t-statistic for γ with $H_0 : \gamma = 0$
	explanatory variables influencing publication bias
Publication year	year of publication, reference year 1990
Publication year^2	square of *year*
Working paper	dummy=1 if study is a working paper, 0 otherwise
Citations	logarithm of number of citations per year since publication (based on Google Scholar)
Impact	Quality of journal based on IDEAS/RePEc Recursive Discounted Impact Factor
OLS method	dummy=1 if model used was OLS, 0 otherwise
	explanatory variables directly influencing effect size
countryUS	dummy=1 if dataset is from US, 0 otherwise
Years	number of years in the dataset
Average year	average year of data source (minimum+maximum)/2
Dependent variable	dummy=1 if dependent variable was productivity 0 for profitability
Labour	dummy=1 if dependent variable was normalized by labour, 0 otherwise
	dummy variables used for grouping
datasourceid	ID of data source
journalid	ID of Journal
authorid	ID of group of authors
kohli	dummy=1 if study was used by Kohli & Devaraj (2003), 0 otherwise
stiroh	dummy=1 if study was used by Stiroh (2005), 0 otherwise

Table 2.8: Explanatory meta-regression analysis

	Mixed-effects multilevel			Clustered OLS		
	Profitability	Productivity	All	Profitability	Productivity	All
prec	-3.650*	5.133**	5.161**	-2.354	4.843**	4.853**
	(-2.32)	(21.95)	(23.72)	(-1.04)	(14.54)	(9.51)
Observations	0.00968**	-0.00950**	-0.00963**	0.00893**	-0.00788**	-0.00781**
	(5.41)	(-13.42)	(-16.67)	(3.36)	(-7.00)	(-5.92)
Years	0.000896*	-0.000187	0.000305	0.000130	-0.000267	0.000150
	(2.19)	(-0.49)	(1.04)	(0.20)	(-0.42)	(0.26)
Labour	-0.0206**	-0.00816	-0.00449	-0.0295*	-0.00957	-0.00217
	(-2.84)	(-0.91)	(-1.29)	(-2.22)	(-0.98)	(-0.48)
Publication year	-0.00305**	0.00883**	0.00670**	-0.00108	0.00712**	0.00610**
	(-3.03)	(12.40)	(21.39)	(-0.73)	(6.61)	(10.74)
Citations	-0.00924**	0.00603**	0.00369**	-0.00356	0.00377*	0.00429**
	(-3.31)	(2.90)	(3.12)	(-1.01)	(2.25)	(2.70)
Working paper	-0.0115	0.0201	0.0357**	-0.00183	0.0315	0.0348**
	(-1.01)	(1.10)	(4.31)	(-0.12)	(1.38)	(2.97)
OLS method	0.0149**	-0.00322	-0.000689	0.0191†	-0.00107	0.000642
	(3.70)	(-1.32)	(-0.32)	(1.78)	(-0.16)	(0.09)
countryUS	0.0175*	0.0118	0.00420	0.0190*	0.00660	0.00711
	(2.13)	(1.53)	(0.92)	(2.10)	(0.73)	(1.26)
Average year	0.00180*	-0.00258**	-0.00259**	0.00115	-0.00243**	-0.00244**
	(2.28)	(-21.99)	(-23.63)	(1.01)	(-14.44)	(-9.43)
Impact factor	0.00181	-0.0162**	-0.456	-0.00218	-0.0102†	-0.00430
	(0.41)	(-3.16)	(-0.75)	(-0.50)	(-1.80)	(-1.53)
Dependent variable			0.0117**			0.00699†
			(4.11)			(1.96)
Constant	1.618**	1.937**	1.448*	1.531**	1.142**	1.145**
	(3.98)	(2.65)	(2.56)	(3.61)	(3.38)	(4.01)
Observations	317	533	850	317	533	850
rmse				2.259	3.839	3.535

Dependent variable: tstat. t statistics in parentheses. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

2.B Studies Included in the Meta-Analysis

Table 2.9: Studies used for meta-analysis

Cite	Year	Estimates	Kohli	Stiroh
Aboal & Tacsir (2015)	2015	6	No	No
Ahn (1999)	1999	2	No	No
Arvanitis & Loukis (2009)	2009	21	No	No
Barua & Lee (1997)	1997	2	No	No
Barua *et al.* (1995)	1995	5	Yes	No
Becchetti *et al.* (2003)	2003	11	No	No
Berndt & Morrison (1995)	1995	20	No	No
Bertschek & Kaiser (2004)	2004	4	No	No
Bertschek *et al.* (2006)	2006	2	No	No
Bloom *et al.* (2012)	2012	24	No	No
Bresnahan *et al.* (2002)	2002	5	No	Yes
Brynjolfsson & Hitt (1995)	1995	10	Yes	Yes
Brynjolfsson & Hitt (1996)	1996	17	Yes	Yes
Brynjolfsson & Hitt (2003)	2003	113	No	Yes
Brynjolfsson & Hitt (1994)	1994	17	No	No
Byrd & Marshall (1997)	1997	4	Yes	No
Campbell (2012)	2012	10	No	No
Caselli & Paterno (2001)	2001	15	No	Yes
Commander *et al.* (2011)	2011	40	No	No
Devaraj & Kohli (2000)	2000	2	Yes	No
Dewan & Min (1997)	1997	8	Yes	Yes
Dewan *et al.* (2007)	2007	12	No	No
Doms *et al.* (2004)	2004	6	No	No
Falk & Seim (1999)	1999	6	No	No
Gargallo-Castel & Galve-Górriz (2007)	2007	3	No	No
Gargallo-Castel & Galve-Górriz (2012)	2012	1	No	No
Gilchrist *et al.* (2001)	2001	18	No	No
Hall *et al.* (2013)	2013	7	No	No
Hempell (2003)	2003	8	No	No
Hempell (2005a)	2005	14	No	Yes
Hempell (2005b)	2005	32	No	No
Hitt & Brynjolfsson (1996)	1996	10	Yes	No
Hitt *et al.* (2002)	2002	6	No	No
Chowdhury (2006)	2006	4	No	No
Chwelos *et al.* (2010)	2010	10	No	No
Kiley (2001)	2001	3	No	Yes
Koski (1999)	1999	9	No	No
Kudyba & Diwan (2002)	2002	15	No	No
Lee & Barua (1999)	1999	5	Yes	Yes
Lee & Perry (2002)	2002	6	Yes	No
Lehr & Lichtenberg (1998)	1998	3	Yes	No
Lehr & Lichtenberg (1999)	1999	27	No	Yes
Li & Richard Ye (1999)	1999	2	Yes	No
Licht & Moch (1999)	1999	4	No	No
Lichtenberg (1996)	1996	6	Yes	Yes
Liu *et al.* (2014)	2014	1	No	No
Luo & Bu (2016)	2016	13	No	No
Mahr & Kretschmer (2010)	2010	17	No	No
Menon & Lee (2000)	2000	1	Yes	No

Continued on Next Page...

Table 2.9 – Continued

Cite	**Year**	**Estimates**	**Kohli**	**Stiroh**
Menon *et al.* (2000)	2000	1	Yes	No
Mitra (2005)	2005	10	No	No
Moshiri & Simpson (2011)	2011	6	No	No
O'Mahony & Peng (2011)	2011	24	No	No
Paton *et al.* (2004)	2004	10	No	No
Prasad & Harker (1997)	1997	8	Yes	No
Rai *et al.* (1996)	1996	10	Yes	No
Rai *et al.* (1997)	1997	4	Yes	No
Siegel (1997)	1997	18	Yes	No
Sircar *et al.* (2000)	2000	13	Yes	No
Steindel (1992)	1992	8	No	Yes
Stiroh (2002a)	2002	12	No	Yes
Stiroh (2002b)	2002	9	No	Yes
Stiroh (2005)	2005	30	No	No
Tam (1998)	1998	24	No	No
Tambe & Hitt (2012)	2012	38	No	No
van der Wiel (2001)	2001	4	No	No
van der Wiel & van Leeuwen (2003)	2003	10	No	No
Wilson (2009)	2009	14	No	No
Wolff (2002)	2002	1	No	Yes
Yorukoglu (1998)	1998	1	No	No
Zwick (2003)	2003	8	No	No

Chapter 3

The Euro's Trade Effect: A Meta-Analysis

Abstract[1] Many studies have estimated the trade effect of the euro, but their results vary greatly. This meta-analysis collects 3,323 estimates of the euro effect along with 28 characteristics of estimation design from almost 60 studies and quantitatively examines the literature. The results show evidence of publication bias, but they also suggest that the bias decreases over time. After correcting for the bias, the meta-analysis shows that the literature is consistent with an effect ranging between 2 and 6 %. The results from Bayesian model averaging, which takes into account model uncertainty, show that the differences among estimates are systematically driven by data sources, data structure, control variables, and estimation techniques. The mean reported estimate of the euro's trade effect conditional on best-practice approach is 3 %, but is not statistically different from zero.

[1]The paper was published in the *Journal of Economic Surveys* [2019, 33(1), pp. 101–124]. I thank the participants of the MAER-Net Conference in Prague 2015, Andrew K. Rose, Tom Stanley and anonymous referee for their helpful comments. I acknowledge support from the Grant Agency of Charles University (grant #280515) and from the Czech Science Foundation (grant #18-02513S). The views expressed in this paper are those of the author and not necessarily those of the Czech National Bank.

3.1 Introduction

Currency unions have been of great interest in academia for almost two decades. This topic has attracted many researchers, particularly following the article by Rose (2000), who found a very strong positive effect of currency unions (approximately 200%) on bilateral trade. This relation has thus been labelled the Rose effect. Many studies have begun to question this result, and some have replicated the original study and questioned whether the effect holds for a newly created European monetary union (EMU) with a new common currency – the euro.

The positive and large effect of a common currency on trade is very attractive to policy makers who aim to emphasize the benefits of currency unions. Nevertheless, estimates of this effect vary widely across studies and often have ambiguous results. An example of such variance in primary research occurs in the most recently published paper by well-known researchers Glick & Rose (2016) and the working paper version of that study Glick & Rose (2015). The idea of the study was to replicate and improve the original paper from 2000 using new data and methods and to compare the eurozone with other currency areas. Research published between 2000 and 2016 suggests that there is a difference between the eurozone and these areas[2] Glick & Rose (2015) find that the EMU is different from other currency unions, that methodology is important, and mainly that the "EMU has a mildly stimulating effect on trade at best". They conclude that "it is currently beyond our ability to estimate the effect of currency unions on aggregate trade with much confidence". Later, Glick & Rose (2016) concluded that the "EMU has thus far boosted bilateral trade by around 50%". This significant change in presented results and the papers as a whole might be a bit surprising since Rose co-authored several studies on currency unions – not only on the euro area but also a meta-analysis of all currency unions (Rose & Stanley, 2005). This case nicely illustrates the possible variation in research results, and this study uses meta-analytic techniques to address the heterogeneity and offer a better understanding of the drivers of these results.

The meta-analysis by Rose & Stanley (2005) was followed by Havránek (2010), who compares the eurozone to other currency areas using only one estimate per study (28 for the eurozone) and focusing only on sources of publi-

[2]This study examines studies focused on the eurozone in terms of the effects on bilateral trade; a list of examined studies is presented in the appendix.

cation bias. The latest meta-analysis by Rose (2016) extends the dataset to 45 studies but continues to use only one estimate per study, which is a substantial limitation and can make the biases even worse, as argued by Viscusi (2017). Additionally, Rose does not address publication bias as this study does.

This paper goes beyond Havránek (2010) and Rose (2016) because it uses over 100 times more estimates of the effect of the euro on bilateral trade from 57 studies spanning two decades of research. This extensive dataset not only allows us to use more advanced and up-to-date methods but also allows us to avoid the subjectivity of researchers picking their preferred estimate. In addition, this paper obtains much more robust results. New evidence on the Rose effect and publication bias is based on 3323 observations, making this the largest meta-analysis in the field of economics to date.

To conduct a proper systematic review of the empirical research, this paper exploits meta-analysis techniques. These methods help us reveal the underlying effect of the euro on bilateral trade and allow us to consider which aspects of research lead to the observed variation in the results. Meta-analysis was introduced to the field of economics by Stanley & Jarrell (1989) and is now widely used in many areas; for example, recent analyses have been conducted in international economics (Havránek & Iršová, 2017) and productivity (Polák, 2016). This paper does not pioneer meta-analytic methods but follows previous meta-analyses focusing on currency unions and on the euro area. The first analysis by Rose & Stanley (2005) finds the effect of currency unions to be between 30 and 90 per cent. Havránek (2010) considers separate effects for the eurozone and other currency unions, making that study the first meta-analysis with a pure focus on the eurozone. His results show that whereas other currency unions boost trade by approximately 60%, there is no effect for the eurozone, and the results are biased upwards by strong publication bias. Subsequently, Rose (2016) selects a single estimate from each study, and the effect size of his preferred specification is 12.3% without controlling for publication bias. Baldwin & Taglioni (2007) notes that the effects of the euro are also biased upwards as a result of the estimation techniques used. Based on these findings, I include more details about the primary studies and their estimation methods in my analysis and follow meta-analytic guidelines (Stanley *et al.*, 2013).

Publication bias has been found in many areas of empirical economics and is proven to be a very serious issue (Stanley, 2005). Publication bias, which stems from the motivation to be published, results in a preference for statistically significant estimates over insignificant ones and for estimates that are in line with theoretical expectations over those that are not. For currency unions, the positive effect of a common currency is expected, and based on evidence presented by Havránek (2010), one needs to account for potential publication bias to avoid exaggerating the effect of the euro on trade. Testing for publication bias is also recommended in meta-analyses.

We contribute to the literature in several ways: first, this work constructs a dataset comprising 57 papers; second, this study includes all the estimated effect sizes reported in the examined papers – Havránek (2010) and Rose (2016) use only one author-preferred estimate from each study; third, this study includes new variables in the explanatory meta-regression analysis with a focus on the methodology; fourth, this study uses new estimation techniques; and fifth, this study accounts for the global financial crisis (GFC).

The remainder of this paper is structured as follows. Section 3.2 presents the theory used to estimate the effect of currency unions and describes the process of selecting studies and collecting data, as well as the properties of the data. Section 3.3 explains the meta-analytic methods I use to analyse the data. Section 3.4 provides and discusses the empirical results. Section 3.5 predicts the estimate of euro based on best practices, and section 4.7 concludes.

3.2 The Rose Effect Dataset

Currency union theory is a subfield of international economics and trade and is commonly examined by gravity models. Gravity models are rooted in physics, where the attraction between two objects is proportional to their size divided by their distance. In trade, we consider countries and their GDP. For more detail on the model specification, see Baldwin (2006). The following is the conventional form of the gravity model used in international trade:

$$\log T_{ijt} = \alpha_0 + \gamma CU_{ijt} + \chi_1 (\log Y_i \cdot \log Y_j) + \chi_2 \log D_{ij} + \chi_3 RTA_{ij} + \sum_{k=1}^{K} \eta_k X_{ijt} + \epsilon_{ijt} \tag{3.1}$$

where T_{ijt} stands for bilateral trade between countries i and j in period t, Y denotes real GDP, and D measures the distance between countries (generally, between their geographical centres). The dummy variable RTA indicates a trade agreement, such as an FTA (Free trade agreement), the EFTA (European Free Trade Area), or the EU, and the dummy variable CU equals one if both countries are in the monetary union and use the same currency at time t. Numerous control variables are captured by vector X, including the use of the same language, having a common border, and the variance of the exchange rate. ϵ is a disturbance term. We are interested in the coefficient γ, which determines the effect of a monetary union on bilateral trade, *ceteris paribus*. The following hypotheses are tested: $H_0 : \gamma = 0, H_A : \gamma \neq 0$. It is important to note at this stage that to obtain the actual effect in percentage terms, one has to calculate the effect as $(\exp \gamma - 1) * 100$ even if we can say that $\lambda \doteq e^{\gamma} - 1$ for small values.

My analysis focuses on the eurozone, which was formally established in 1999 by 11 EU member states. The euro has been a circulating currency since 2002, and 19 EU member states currently use the euro. The primary research has evaluated only the ex post effects of joining the EMU on trade flows. Since the EMU began as a political impulse, the results are used to emphasize the positive effects of the single currency. Doucouliagos & Stanley (2013) prove that the desire to find a positive effect – and possibly a larger effect – is an important driver of publication bias, which, in this case, can greatly overstate the true effect size.

The presented gravity equation is still valid in its general terms, but there has been a great deal of development regarding estimation techniques, the control variables used and the type of data that influence the model results. For a more detailed survey and comparison of methods, see Head & Mayer (2014); for a multilateral resistance problem, see Anderson & Van Wincoop (2003); and for medal mistakes, see Baldwin & Taglioni (2007).

Table 3.1 shows the mean estimates for several groups. Since this study uses all estimates from each paper, the right-hand side of the table shows estimates weighted by the inverse of the number of estimates reported in the study. These weights mean that we treat each study equally and do not allow the results be affected more by studies with higher numbers of reported estimates. The mean unweighted estimate of semi-elasticity is 0.09, which implies that the euro effect is $e^{0.09} - 1 = 9.4\%$. This relation means that on average, the euro boosted trade between the countries using it by 9.4 percent, and the 95% confidence interval for the mean estimate ranges from 8.3 percent to 11.4 percent, which is not very substantial at first glance. The table indicates that the semi-elasticities vary substantially when categorized by basic characteristics. The largest difference is between the mean of the estimates of the main results and the mean of the estimates of the robustness checks, which is much higher. The table also suggests that the reported euro effect size estimates are getting smaller (if we use Havránek (2010) as a breaking point). This result might be caused by the evolution of methods or the study designs influencing the estimates. In all categories, the weighted mean is smaller that the simple average. This study uses meta-regression analysis to investigate the marginal effects of data and method choices on the reported euro effects in Section 3.3.

Table 3.1: Euro effects for subsets of studies

	No. of estimates	Median	Mean	Weighted mean
All estimates	3323	0.08	0.09	0.08
Estimates before Havránek (2010)	1581	0.1	0.13	0.04
Estimates after Havránek (2010)	1742	0.06	0.07	0.04
Non-published	2280	0.08	0.09	0.04
Published	1043	0.09	0.1	0.05
Main results	659	0.08	0.04	0.03
Robustness check	2664	0.08	0.11	0.05
IMF data	548	0.1	0.06	0.06
Other data	2775	0.08	0.1	0.03

Notes: The table presents mean estimates of the euro coefficient (the coefficient estimated in a gravity equation on the dummy variable that is equal one if both trade partners are part of the eurozone) for estimates reported in a particular study. Estimates before/after Havránek (2010) = estimates from studies analysed/non-analysed by Havránek (2010). Main results = estimates that are said to be the main results of each study. Robustness check = estimates from results labelled as robustness check or part of the appendix. IMF data = estimates from studies that use the IMF/DOTS data source. Other data = estimates from studies that used data sources other than IMF/DOTS. For the weights, the inverse of the number of the reported estimates in a particular study was used. Values are rounded to the closest hundredth.

Figure 4.2 shows the histogram of the estimated euro semi-elasticities. We observe that a large proportion of estimates are very close to zero. The overall mean is equal to 0.09, which is almost the same as the weighted mean, and the shape of the distribution suggests that there are no serious coding mistakes or outliers in the dataset that should be excluded from the analysis.

The data for this study are collected from all studies on the effect of the euro on trade. I began with the studies analysed by Havránek (2010) and Rose (2008) and then searched the Google Scholar and RePEC databases for both published and unpublished studies using the following keywords: *euro, trade, EMU, effect, Rose*. When I found new, updated or published versions of any working papers, I included them in my dataset. Of all the studies

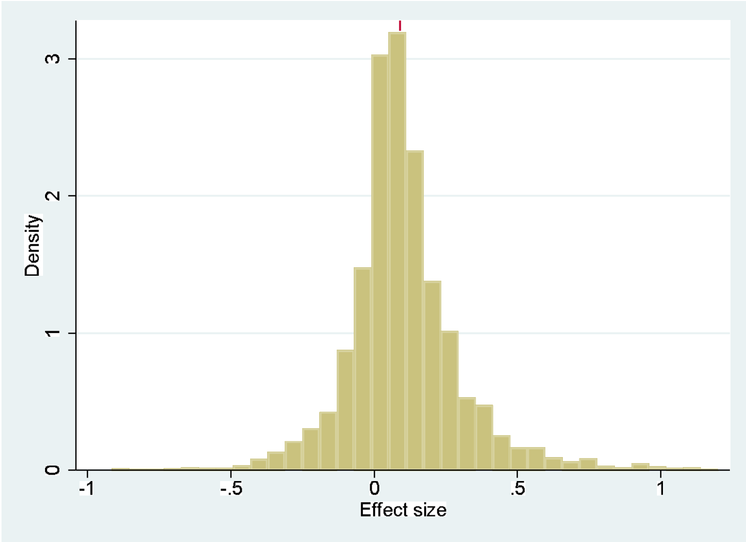


Figure 3.1: Euro estimates are evenly distributed around the mean

Notes: The figure shows the histogram of the estimates of the euro coefficient (the coefficient estimated in a gravity equation on the dummy variable that is equal one if both trade partners are part of the eurozone) reported in individual studies. The solid vertical line denotes the mean of all estimates.

identified, I take into account only those for which the empirical framework uses gravity equations and those that focus on the euro area. These conditions result in a dataset of 57 studies – more than double the number included in the meta-analysis by Havránek (2010) – and 3323 estimates, which is over 100 times greater than the 28 estimates used by previous studies. The current best practice in meta-analysis is to use all available estimates rather than one author-preferred estimate per study, which is not only much more transparent but also allows for the use of more control variables. Compared with previous works, i.e., Havránek (2010) and Rose (2016), my dataset is more focused on methods and data characteristics and less focused on researcher characteristics. The list of collected variables is based on those gathered in previous international trade research and on best practices for accounting for publication bias in meta-analysis.

Table 4.1 lists all the variables collected from primary studies and provides variable definitions and basic summary statistics. To improve understanding, the 28 collected variables are classified into groups – data characteristics, estimation methods, control variables and publication characteristics. The intention of this procedure is to examine possible sources of heterogeneity in the estimated euro effect, to provide greater insight into any differences in the results and to address the most well-known issues that are said to drive the results of gravity model estimations. These variables should cover the choices made by researchers and the methodological issues analysed by Baldwin & Taglioni (2007) and Head & Mayer (2014).

Table 3.2: Description of regression variables

Variable	Description	Mean	SD
Euro	The coefficient estimated in a gravity equation on the dummy variable that equals one if the trading county pair is part of euro zone area	0.1	0.21
SE	Estimated standard error of *Euro*.	0.19	0.50
Precision	Inverse of the estimated standard error of *Euro*.		
Data characteristics			
Midyear of data	Midpoint of the sample on which the gravity equation is estimated.	1998	4.70
Disaggregated	= 1 if trade flows are disaggregated at the sector or product level	0.55	0.50
Data IMF	= 1 if data source if IMF/DOTS, 0 otherwise	0.16	0.37
Robustness check	= 1 if coefficient is part of appendix or robustness check	0.80	0.40
Dependent variable[3]	= 1 if dependent variable is based on import only	0.20	0.40
No. of years	Number of years in the data (timespan of the dataset in the primary study).	16.66	8.60
Years in EMU	Number of years in the data since the adoption of euro.	6.49	3.47
Obs. per year	The logarithm of the number of observations per year included in the gravity equation.	6.24	2.20
Crisis	=1 if dataset covers years from 2009 onwards where trade was affected by the global financial crisis.	0.21	0.41
EU data only	=1 if dataset covers only EU countries.	0.20	0.40
Total trade	= 1 if total trade is used as the dependent variable and imports and exports are summed before taking logs.	0.21	0.41
Countries	Number of countries in the dataset	30.27	37.65

Continued on next page

[3]Based on a reviewer suggestion, we also coded if the variable is base on one-way trade only, but such variable was omitted in the regression due to multicollinearity, therefore we excluded it.

Table 3.2: Description of regression variables (continued)

Variable	Description	Mean	SD
Treatment of multilateral resistance and estimation methods			
PPML	= 1 if the gravity equation is estimated by the Poisson pseudo-maximum likelihood estimator	0.01	0.11
Time-varying fixed eff.	=1 if the gravity equation is estimated using time-varying fixed effects.	0.18	0.39
No control for MR	=1 if gravity equation does not account for multilateral resistance	0.10	0.30
Anderson and Wincoop	=1 if study discusses multilateral resistance or refers to Anderson & Van Wincoop (2003).	0.84	0.37
Control variables			
FTA control	= 1 if the gravity equation controls for free trade agreements when possible.	0.86	0.35
EU dummy	= 1 if dummy variable for EU is used in the gravity equation.	0.54	0.50
Missing control var.	= 1 if gravity equation does not control for adjacency, distance or shared language but it is possible.	0.19	0.39
Real exchange rate	= 1 if real exchange rate is used in the gravity equation.	0.46	0.50
Volatility	= 1 if volatility of exchange rate is used in the gravity equation.	0.35	0.48
Publication characteristics			
Working paper	= 1 if the study is not published in a peer-reviewed journal.	0.69	0.46
Impact	Recursive discounted RePEc impact factor of the outlet (collected in September 2017).	0.21	0.42
Citations	The logarithm of the mean number of Google Scholar citations received per year since the study appeared in Google Scholar (collected in September 2017).	1.14	1.07
Publication year	Year when the study first appeared in Google Scholar.	2008.11	3.79
Havránek	= 1 if the study is examined by Havránek (2010)	0.48	0.50

Notes: SD = standard deviation. All variables except for citations and the impact factor are collected from studies estimating the euro effect (the search for studies was terminated on September 1, 2017, and the list of studies is available in Appendix). Citations are collected from Google Scholar, and the impact factor is collected from RePEc. I also tried to control for transparency of authors – whether data or code for the primary analysis are available online. Only Rose (2016) published data for their study. All estimates are from regressions using panel data; therefore, we do not have dummy variables for non-panel data estimates. Values are rounded to the closest hundredth.

Data characteristics The results of models are always driven by the underlying data. If there is heterogeneity between the estimates of the euro effect, data source characteristics need to be controlled for. This study controls for the level of aggregation because, as Anderson & Van Wincoop (2004) indicate, if data are aggregated at the sector or country level, the gravity equation yields different results. The average of estimates based on aggregated data is smaller by 0.01 in the analysed dataset. A larger amount of data should provide more precise estimates; therefore, the logarithm of the number of observations per year is included as an explanatory variable. Baldwin & Taglioni (2007) described several mistakes that researches commit when estimating the gravity equation. One of these mistakes is called the "silver medal mistake", which means that the total trade flow is taken as the response variable but that the sum or average is calculated before taking the log of this value. Following Head & Mayer (2014) and Rose (2016), we argue that up-to-date models are based on (log) exports rather than trade. This study therefore accounts for the type of trade that is used (exports or imports). Furthermore, this meta-analysis controls for the GFC (which occurred in 2008 and 2009) in the primary study dataset because it caused an overall economic downturn and related decline in trading activity.

Treatment of multilateral resistance The topic of multilateral resistance has been highlighted by Anderson & Van Wincoop (2003). A dummy variable is used to investigate whether estimates from studies that reference or discuss multilateral resistance differ from those that do not. The easiest remedy for multilateral resistance is to use fixed effects for the destination and origin country. Since most studies use fixed effects, this study codes studies that use no remedy for multilateral resistance and those that use the theory-consistent estimation technique: time-varying fixed effects. Another dummy variable groups estimates that are based on a Poisson pseudo-maximum likelihood estimator.

Control variables Control variables are an inseparable component of gravity equations. Typically, dummy variables for common language, common borders and free trade agreement membership are included. The next control variable to be included is distance, which is also time-invariant and the same for a given country pair. In many cases, these control variables cannot be included in the regression when the euro semi-elasticity is estimated either because they would be same for all the country pairs (EU countries are part of the same free trade area) or because a specific fixed effects method is used. When a control variable can be included but is omitted, this study codes this feature by setting "Missing control variables" to 1. Another control variable captures the usage and volatility of the exchange rate and real exchange rate in the model. In particular, controlling for the EU dummy variable is important in the context of measuring the effect of euro currency and not of deeper integration within the EU since all eurozone members are also members of the EU.

Publication characteristics Publication bias is a topic that must be examined in meta-analysis. To determine whether published studies differ from working papers and unpublished studies, the dummy variable "Working paper" is employed and is equal to one if a study is not published in a peer-reviewed journal. To account for differences in journal quality, this study follows Zigraiova & Havranek (2016); Havranek (2015); Havranek *et al.* (2015) and uses the recursive discounted RePEc impact factor, which also provides information on working paper series. To capture any opinion-related time trends, the publication year of each study and the number of citations in Google Scholar are coded. To avoid discriminating against the most recent studies, a logarithm of the mean number of citations per year is used in the analysis.

Publication bias Meta-analysis has become one of the leading streams that highlights the problems caused by publication bias. Publication bias has been found in many areas of academic literature, including area of currency unions and their impact on trade as Rose & Stanley (2005) and Havránek (2010). This study aims to examine, if the publication bias is still present in this area of literature of whether it has diminished and authors do not disregard negative results and there is less estimate selection.

In the introduction, I indicated that the focus of this study was on the methodology and study design aspects of papers, mainly controlling for multilateral resistance and using total trade, which are both important issues highlighted by Baldwin & Taglioni (2007). It is quite interesting that only a few researchers discuss this issue in their studies, even if they were published after that date. Fortunately, one of the remedies for the so-called "gold medal mistake" is to use country fixed effects, and a large portion of the estimates were obtained using this method. The second issue is related to the incorrect handling of the dependent variable in the gravity equation, which is the logarithm of trade. Very often (i.e., in approximately one-quarter of the cases), the sum of imports and exports is taken before the logarithm. Making this mistake increases the size of the estimates (variable *Total trade*). Using import data instead of export data should lead to bias as well. Import data are likely to be underestimated because there is an intention to hide some imports to avoid paying import taxes. Export data are therefore more precise. My results for *Dependent variable* indicate a stronger effect of the euro on imports than on exports. This result could mean that within the eurozone, information about imports is less biased than the information about imports reported by other countries.

The variables described here that are intended to explain the heterogeneity in the gathered euro semi-elasticities offer a large number of combinations that can be used in explanatory regressions. Some of these variables would also be redundant in such a regression. Because this fact is not known a priori and should included – and to maintain the least biased method for decision making and following recent trends in meta-analysis, this study uses a Bayesian model averaging (BMA) methodology to address model uncertainty[4].

BMA is ideal for running regressions on different sub-samples of the data to see which results are the most robust. Estimating these regressions would, of course, take a very long time; therefore, I use the bms R package from Feldkircher & Zeugner (2009). This package relies on an algorithm that estimates potential models. More general details on BMA are presented in other studies, e.g., Eicher *et al.* (2011), and the key results are as follows. For each model, BMA calculates the posterior model probability (PMP), which is an information criterion about how well the model fits the data. The reported regression coefficients are weighted averages of the estimated models, and as weights, the PMP are used. To obtain the precision of the coefficients, BMA does not calculate standard errors but rather posterior standard deviations, which are based on the distribution of coefficients from the estimated models. Then, BMA reports for each coefficient the posterior inclusion probability (PIP), which is the sum of the PMP in which the variable is included. PIP simply indicates how likely this variable is to be included in the true model.

3.3 Methodology

The most common method for testing for publication bias in meta-analyses is the funnel asymmetry test (FAT) (Stanley & Doucouliagos, 2010). This approach plots the effect sizes on the horizontal axis and the precision of the estimate – generally, the inverse of the standard error – on the vertical axis.

[4]BMA has been used in meta-analysis, among others, Iršová & Havránek (2013), Havránek *et al.* (2017), Havránek & Iršová (2017), Havranek *et al.* (2018a), and Havranek *et al.* (2018b)

The most precise estimates should be close to the true effect size, and less precise estimates would be farther apart[5]. This method should produce a plot that looks like an inverted funnel. Using a large number of observations, one would expect that the estimates would be normally distributed around the 'true effect'. The assumption of normally distributed estimated semi-elasticities is standard in meta-analyses in the absence of publication bias (Stanley, 2008). When there is no publication bias, the funnel is symmetric since any effect size can be reported. However, this technique provides only a visualisation of the collected data; any conclusions about the presence of publication bias require specialized regression methods. For a currency union, a positive effect is expected, and (Havránek, 2010) previously found that funnel plots for currency unions are "missing" negative observations. Using econometric methods to analyse the funnel plot – FAT-PET – meta-regression analysis can be employed to identify the signal of a non-zero true effect should one exist in the research record[6]. Both approaches are empirically tested using model 3.4.

The funnel plot is a graphical representation of the effect size and the relation between this effect size and its precision. The methods used to estimate semi-elasticity yield a symmetrical distribution; hence, the semi-elasticity values and standard errors should be independent from a statistical perspective (Stanley & Doucouliagos, 2010). If researchers prefer statistically significant results, they seek either large effects or high precision if the effect size is near zero. Either way, this approach leads to correlated semi-elasticities and standard errors. To test for funnel asymmetry and publication bias, I follow Havránek *et al.* (2012); Stanley *et al.* (2008) and use the following equation:

$$\gamma_{ij} = \gamma_0 + \beta \cdot SE\left(\gamma_{ij}\right) + \epsilon_{ij} \tag{3.2}$$

where γ_{ij} is the i-th semi-elasticity estimate of the euro effect from the j-th study, $SE\left(\gamma_{ij}\right)$ is the reported standard error of this estimate, γ_0 is the mean semi-elasticity corrected for publication bias, β is the measure of the publication bias, and ϵ_{ij} is a normal disturbance term. Havránek & Iršová (2017) note that the funnel asymmetry test using equation 4.7 has low power if the true effect is close to zero and that the only source of publication bias is

[5]This assumes, of course, that some estimates in this area of research are drawn from a sampling distribution that has this 'true effect' as its mean.

[6]There is a limitation stated by Alinaghi & Reed (2018) and later expressed by Stanley & Doucouliagos (2019): All meta-analysis methods fail to distinguish a genuine effect from the artefact of publication bias reliably under common conditions found in economics research.

statistically significant. In the literature on the euro effect, publication bias is based on sign rather than significance since economic theory predicts a positive effect and because it is difficult to explain the negative effects of joining a monetary union.

In this study, I also explore the sources of heterogeneity across studies. The standard meta-regression model uses a set of explanatory variables (X) to capture and describe the diverse results found in primary studies. I can add this set of variables to model 4.7 to obtain model 3.3:

$$\gamma_{ij} = \gamma_0 + \beta \cdot SE\left(\gamma_{ij}\right) + \sum_{k=1}^{K} \alpha_k X_{ijk} + \epsilon_{ij} \tag{3.3}$$

where γ_{ij} is the i-th semi-elasticity estimate of the euro effect of the j-th study, X_{ijk} is an independent variable that measures the characteristics of the primary study that lead to the diversity of results, and α_k measures the effects of such characteristics on the estimate of interest. X_{ijk} is commonly a dummy variable, as in my dataset (see Table 4.1).

However, models 4.7 and 3.3 suffer from heteroskedasticity – the variance of the estimated coefficients in the literature is diverse since the primary studies differ in terms of their data sources, sample sizes, independent variable selection and estimation methods. To cope with heteroskedastic errors in the meta-regression, I use the weighted least squares (WLS) methodology presented by Stanley (2005), which is recommended by Stanley *et al.* (2008) and followed by, e.g., Havránek *et al.* (2012). In the WLS regression, I use the standard errors as weights and add study-level effects. I divide the regression equations 4.7 and 3.3 by the estimated standard errors and obtain models 3.4 and 3.5, respectively, with the t-statistic as the dependent variable:

$$\frac{\gamma_{ij}}{SE\left(\gamma_{ij}\right)} = t_{ij} = \gamma_0 \cdot \frac{1}{SE\left(\gamma_{ij}\right)} + \beta + e_{ij} \tag{3.4}$$

and:

$$t_{ij} = \gamma_0 \cdot \frac{1}{SE\left(\gamma_{ij}\right)} + \beta + \sum_{k=1}^{K} \alpha_k \frac{X_{ijk}}{SE(\gamma_{ij})} + e_{ij} \tag{3.5}$$

After this modification, the interpretation of the coefficients in equation 3.4 is the same: γ_0 is the mean semi-elasticity corrected for publication bias, β is the measure of publication bias, and e_i is the disturbance term. Cipollina & Salvatici (2010) emphasizes that regression 3.5 may still lead to consistent yet inefficient estimators since the estimates from the same study j are not independent. As a remedy, a clustering procedure is undertaken to adjust the standard errors for intra-study correlation. Each study is taken as a cluster, and the variance-covariance matrix is adjusted accordingly, which is necessary when using OLS.

3.4 Results

3.4.1 Publication bias

Figure 3.2 depicts the funnel plot for all semi-elasticities in my dataset (approximately 20 observations are not shown since the ranges on the axes are restricted to capture the main distribution). The funnel looks quite symmetric compared to the funnel plot provided by Havránek (2010), which was highly asymmetric and included almost no negative estimates. However, Havránek (2010) used only one researcher-preferred estimate per study, whereas this paper uses all available estimates.

The funnel plot helps illustrate the publication bias present; however, empirical tests are necessary to reveal the true magnitude. Table 3.3 provides the results of the regression based on model 3.4, when all estimates are used. I estimate WLS regressions with clustered errors, fixed effects, random effects, and an instrumental variable: the inverse of the logarithm of the number of observations per year. In meta-analysis, instrumental variables that are linked to the number of observations are used. The reason for using these instruments is to correct for error-in-variable bias (Davidson & MacKinnon, 2004). As Stanley (2005) argues, the $\sqrt{n}$ is the obvious instrument to use for $1/SE_i$. Therefore, the most commonly used instrument is srqt(n) since a greater number of observations used to estimate the coefficient yields more precise estimates; however, there is no linear relationship. This paper uses a slight modification of this value to obtain a higher correlation between the standard error and the instrument (0.2). The instrument, when constructed in the usual way, provides almost identical results but is less suitable from an econometric perspective.

As a robustness check of the effect size, I estimate the precision effect estimate with standard errors (PEESE). PEESE should always be employed if one rejects $H_0 : \gamma = 0$ because it provides a better estimate of the genuine effect if there is some evidence of its existence. The limitation of PEESE is rapidly reduced if publication bias or strong heterogeneity is present (σ^2 reaches high values). Additionally, PEESE has almost no power if the sample

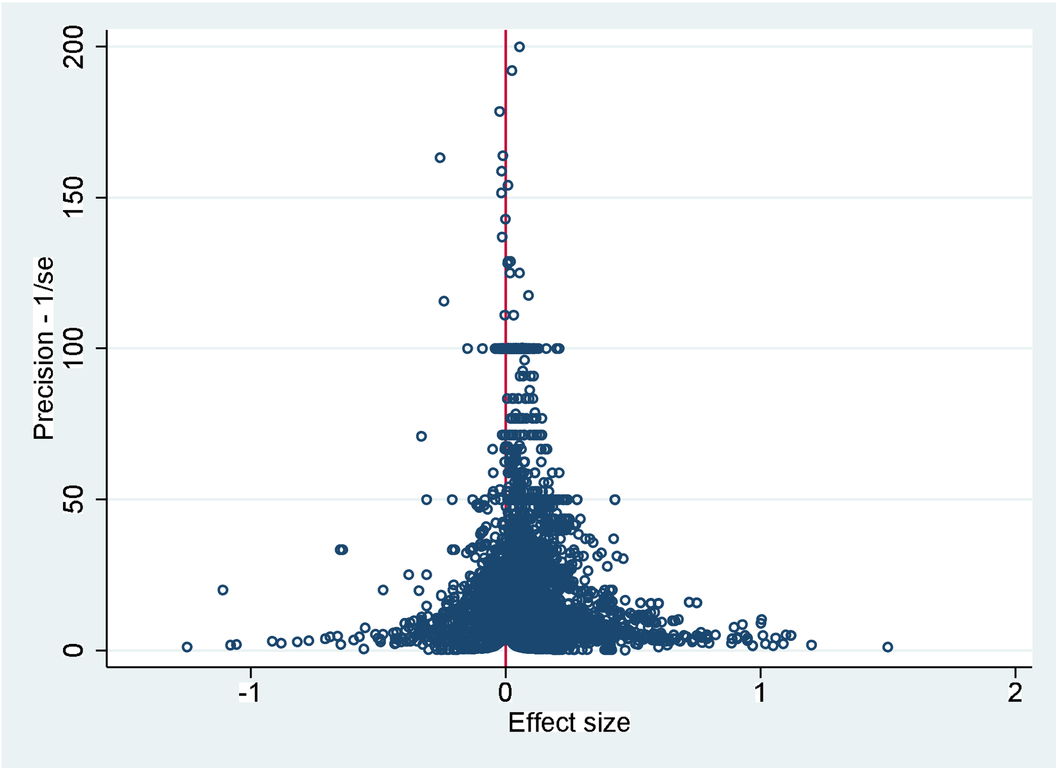


Figure 3.2: Funnel plot – all estimates from all studies

Notes: The funnel plot is a scatter plot where estimate (Effect size) is plotted against its precision (inverse of standard error). In the absence of publication bias the funnel should be symmetrical around the value given by estimates with the highest precision. The vertical line denotes the zero value, mean and median values are reported in Table 3.1 .

for meta-analysis is constructed using only small studies, which is not case for this study (Stanley, 2005; Moreno *et al.*, 2009; Stanley, 2017). PEESE is tested using Equation 3.6, following Equation 3.4, where γ_0 is the estimate of euro semi-elasticity corrected for publication bias and β measures the bias:

$$\frac{\gamma_{ij}}{SE\left(\gamma_{ij}\right)} = t_{ij} = \gamma_0 \cdot \frac{1}{SE\left(\gamma_{ij}\right)} + \beta \cdot SE\left(\gamma_{ij}\right) + e_{ij} \tag{3.6}$$

Table 3.3: FAT-PET-PEESE – whole dataset

	WLS	PEESE	IV	FE	RE
Precision - 1/SE	0.0349^{**}	0.0539^{**}	0.0398^{**}	$0.0157^{\dagger}$	0.0159^{**}
(Effect beyond bias)	(5.62)	(33.51)	(4.17)	(1.71)	(5.54)
SE		0.163			
(Publication bias)		(1.62)			
Constant	0.949^{**}		0.902^{*}	1.390^{**}	1.798^{**}
(Publication bias)	(9.13)		(2.52)	(6.62)	(5.50)
Observations	3323	3323	3323	3323	3323
Studies	57	57	57	57	57

Notes: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; t statistics in parentheses; the table presents the results of regressions 3.4 and 3.6; dependent variable: tstat of i-th estimate of euro coefficient; standard errors clustered by study; FE = fixed effects; RE = random effects; IV = instrumental variable, instrument is the inverse of logarithm of the number of observations per year.

The results presented in Table 3.3 confirm the graphical test and provide two new pieces of evidence. First, there is a positive and significant effect of the euro on bilateral trade between countries, but this effect is estimated to be between 1.5 % and 5.5 %, which is approximately ten times smaller than the preferred estimate in the most recent study by Glick & Rose (2016). However, the recent meta-analysis by Havránek (2010) on the euro effect found no significant effect at all. Second, the magnitude of publication bias is weaker than that previously found by Havránek (2010) and Rose & Stanley (2005). The bias value does not reach 2, which would be evidence of strong publication bias (Doucouliagos & Stanley, 2013).

New evidence indicates a change in publication bias in the literature focusing on the effect of the euro on bilateral trade. For a deeper analysis, I split the sample into two parts in the next step – the studies included in Havránek (2010) and newer studies (identified in Table 4.6). The re-drawn funnel plot capturing the two subsamples is depicted in Figure 3.3 (for more detailed charts, see Figures 3.5 and 3.6 in the appendix). The funnel plot of previously examined studies (the plot on the right-hand side) is skewed – the funnel is not very symmetrical; negative and less precise estimates are "missing". For studies published after Havránek (2010), (the plot on the left-hand side) the funnel is much more symmetric. Publication bias in this area of the literature is probably smaller, but to be sure, I conduct the necessary empirical estimation of regression model 3.4 for these samples. The empirical FAT-PET-PEESE results are provided in Tables 3.4 and 3.5. Accordingly, I follow a WLS approach with clustered standard errors, an instrumental variable and fixed effects.

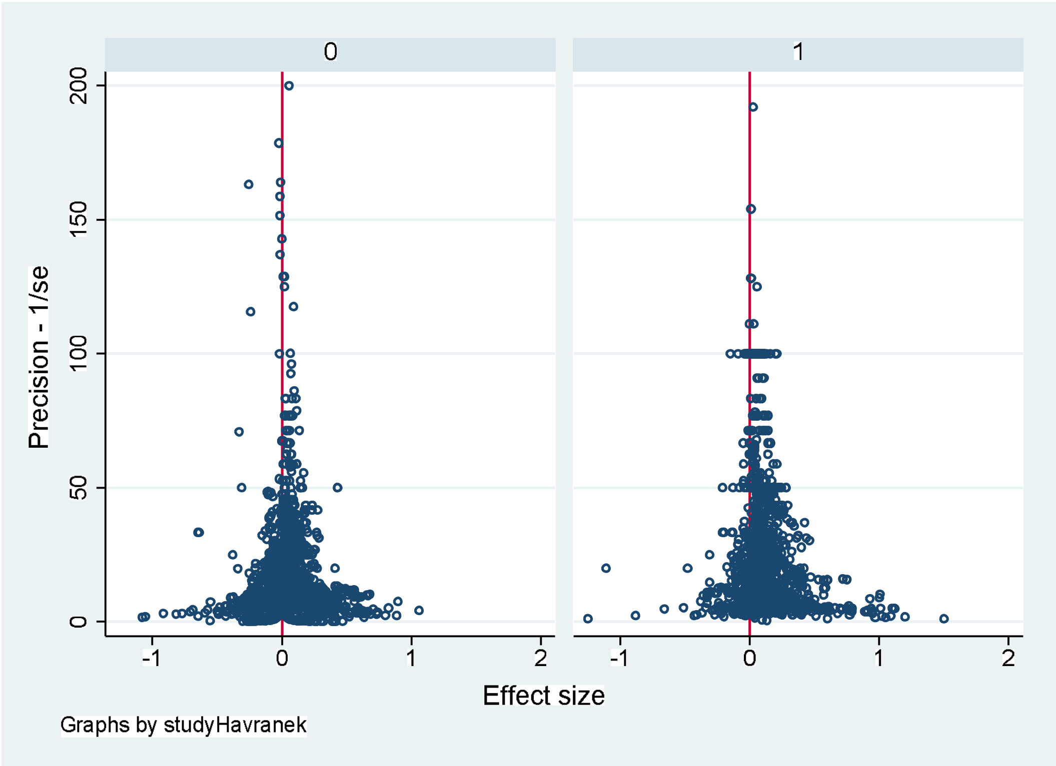


Figure 3.3: Funnel plot – new studies compared to those examined by Havránek (2010)

Notes: The funnel plot is a scatter plot where estimate (Effect size) is plotted against its precision (inverse of standard error). Left-hand side of figure depicts all estimates from primary studies published after Havránek (2010), right-hand side of the figure depicts all estimates from studies included by Havránek (2010). In the absence of publication bias the funnel should be symmetrical around the value given by estimates with the highest precision. The vertical line denotes the zero value, mean and median values are reported in Table 3.1 The right-hand side of the figure suggest presence of publication bias caused by "missing" negative estimates.

The results in Tables 3.4 and 3.5 confirm my hypothesis. Publication bias is clearly present in the studies examined by Havránek (2010), as confirmed by the results in Table 3.4. The estimated effect size of the euro is positive and significant. For newly examined studies, we observe a symmetric funnel

Table 3.4: FAT-PET-PEESE – studies examined by Havránek (2010)

	WLS	PEESE	IV	FE
Precision - 1/se	0.0339**	0.0614**	0.0393**	0.0270*
(Effect beyond bias)	(4.25)	(30.21)	(3.97)	(2.48)
Standard error		4.107**		
(Publication bias)		(6.86)		
Constant	1.699**		1.585**	1.905**
(Publication bias)	(9.06)		(3.20)	(5.85)
Observations	1581	1581	1581	1581
Studies	26	26	26	26

Notes: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; t statistics in parentheses; the table presents the results of regressions 3.4 and 3.6; dependent variable: tstat of i-th estimate of euro coefficient; standard errors clustered by study; FE = fixed effects; IV = instrumental variable, instrument is the inverse of logarithm of the number of observations per year.

Table 3.5: FAT-PET-PEESE – new studies after Havránek (2010)

	WLS	PEESE	IV	FE
Precision - 1/se	0.0133	0.0307**	0.0225	-0.00397
(Effect beyond bias)	(0.97)	(11.27)	(1.53)	(-0.27)
Standard error		0.0626		
(Publication bias)		(0.67)		
Constant	0.659**		0.567†	0.944**
(Publication bias)	(3.50)		(1.94)	(3.86)
Observations	1742	1742	1742	1742
Studies	31	31	31	31

Notes: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; t statistics in parentheses; the table presents the results of regressions 3.4 and 3.6; dependent variable: tstat of i-th estimate of euro coefficient; standard errors clustered by study; FE = fixed effects; IV = instrumental variable, instrument is the inverse of logarithm of the number of observations per year.

plot, and empirical results confirm that there is no strong publication bias. Regarding the effect size of the euro, three methods found no significant values. The PEESE test found a significant euro effect, but as argued, This test should be used only if we reject the null hypothesis and there is small variance. Using the new sample, the results indicate that there is no significant effect of the euro on bilateral trade. If we calculate the effect of the currency in percentages by taking $e^{\gamma}-1$ and re-doing the FAT-PET-PEESE for a robustness check, we obtain almost identical results (see Tables 3.8, 3.9 and 3.10 for more details) and the same conclusion.

3.4.2 Heterogeneity between studies

The preferred methodology for meta-regression analysis is the BMA, which allows us to estimate the regression with unbalanced panels (because studies have different numbers of observations) and to work with a large number of regressors at the same time. The numerical results of the BMA are reported in Table 4.4 along with the results of the WLS regression from equation 3.5, which serves as a robustness check. The WLS estimation method provides results consistent with the BMA exercise since the estimated signs of significant variables are the same and statistically significant variables in the WLS estimation have high PIP, and vice versa. The estimated magnitudes are also almost the same for all significant explanatory variables.

Figure 4.6 graphically captures the results of the BMA exercise. The rows show the individual variables sorted by the the PIP in descending order, meaning that the most important variables are at the top. The columns show different regression models, and the width of the column represents PMP. Models with higher PMP are on the right-hand side. The cells represent the coefficients in the regression. If the cell is white, the variable is not included in the model. Red shading (the lighter shade in grayscale) indicates that the variable is included and that the estimated regression coefficient is negative. If the colour of the cell is blue, the variable is also included, but the estimated regression coefficient has a positive sign. The figure indicates the robustness of the estimated regression parameters – whether the estimated sign varies or not between different models. In our case, the results are robust concerning the estimated sign of the included explanatory variables.

When interpreting the BMA results, this study follows Eicher *et al.* (2011), who consider the value of PIP to be decisive if it exceeds 0.99, strong if it is between 0.95 and 0.99, substantial if it is between 0.75 and 0.95, and weak if it is between 0.5 and 0.75. Because all explanatory variables are divided by standard error, which causes higher multicollinearity, we should also be careful when interpreting the results since these are marginal effects of regressors on the overall effect. Ceteris paribus, in the following sections, we will examine these effects.

Data characteristics The results suggest that data characteristics are an important driver of differences between estimated euro effects. Disaggregated data, IMF/DOTS data sources and a large number of observations in the dataset result in a smaller euro effect. Crisis data also have a negative effect on the euro effect. However, this study does not find a significant difference between the values reported in the main part of the studies compared to the robustness checks and appendices. Furthermore, it seems that the "silver medal mistake" in estimation ($Totaltrade$) does not affect the euro effect, which is the same conclusion Havránek & Iršová (2017) finds for the border effect.

Treatment of multilateral resistance Using up-to-date and proper methods of time-varying fixed effects results in significantly different estimates of the euro effect. This method finds higher estimates than others. Interestingly, results of studies that do not discuss multilateral resistance do not differ from those that do so. The "gold medal mistake" (no control for multilateral resistance) in the gravity equation estimation does not influence the results for the euro effect probably as a result of a small sample, as only 10 % of estimates are based on such regressions.

Control variables This study finds that usage of the FTA dummy has a significant effect on the resulting estimates, and when used, the estimate is smaller. This makes perfect sense because the currency dummy also captures effect of trade agreements that are present in every currency union. Another important control variable is the real exchange rate. Studies that omit this variable report higher estimates than those that do not. Missing control variables (for common language, borders, or distance) results in smaller estimates, which is a bit counter-intuitive because one would expect that the effect of factors such as common borders would be captured by the currency dummy.

Publication characteristics The last group of study characteristics is related to the publication itself. We find differences between working papers and published papers, with the former having a negative effect of the reported size of the euro estimate. This is in line with notion of publication bias, where results that do not align with expectations are supposedly "put in the drawer". Journal quality has a negative impact on the reported effect which means, that results published in high quality journals are smaller than those in lower quality journals, but the number of citations has a positive effect making the bigger estimates more citated. The latter is likely caused by the attractiveness of the larger estimates of the euro effect.

Publication bias The previous section of this study devoted to publication bias presented evidence of publication bias presence in the examined area of academic literature. Results presented in Tables 3.4 and 3.5 suggest, that the publication bias is decreasing over time. To examine this in the MRA, the publication year regressor is interacted with SE. This allows to test for for the decline/increase in publication bias directly. The coefficient is negative, which is the evidence of a decline in publication bias. This finding is also supported by the negative effect of the variable "Havránek" that groups the estimates gathered from the studies examined by Havránek (2010), which were published before 2010. It should be noted, that all moderators are divided by SE, so the study-level effects reflect differential publication bias, which might obscure any trend in publication bias over time.

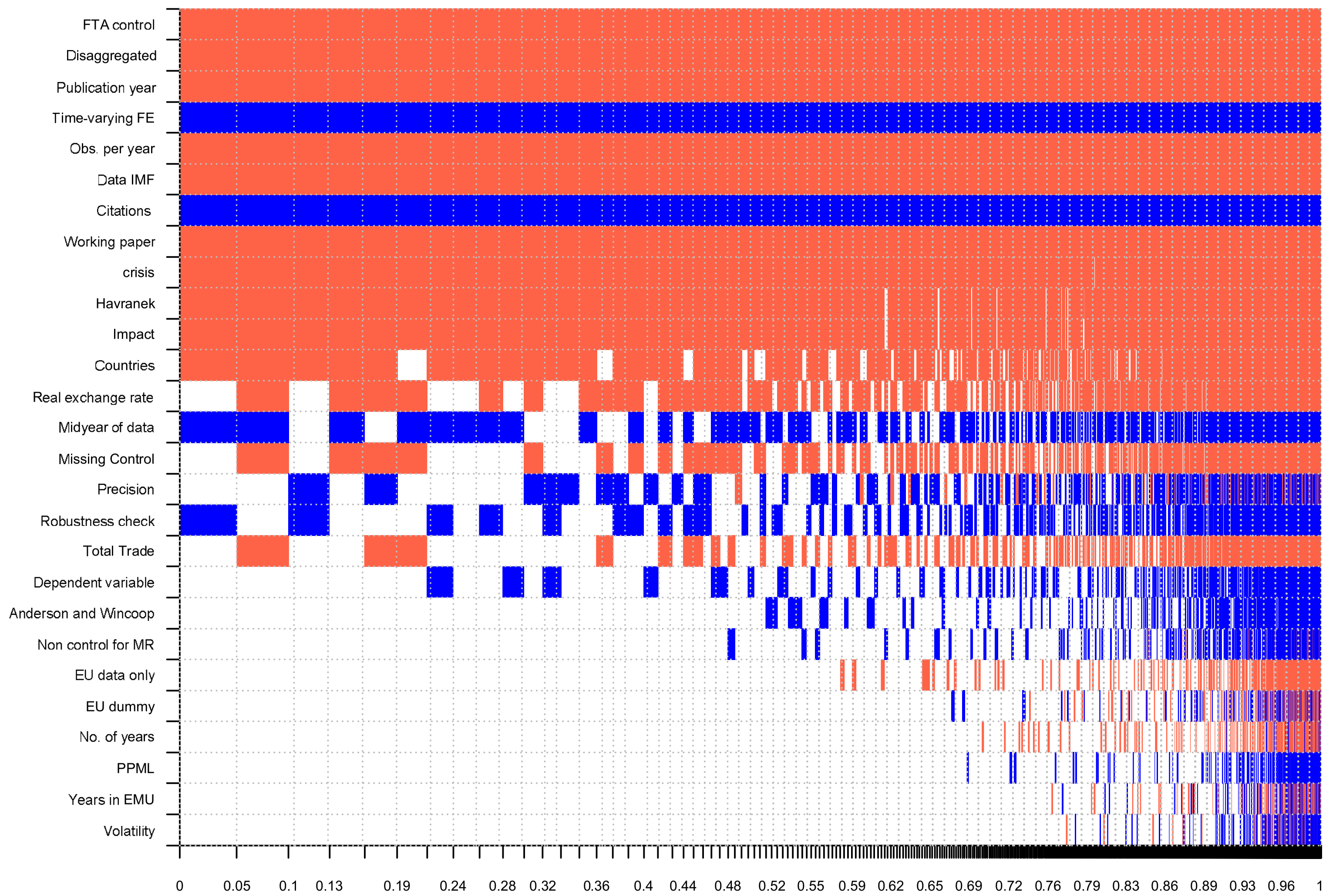
None of the previous studies conducted a meta-regression analysis of studies on the euro only, which makes it impossible to directly compare my results, provided in Table 4.4, to those of previous studies. To some extent, it is possible to compare the effect sizes of a few variables with Havránek (2010, Tab. 3). Consistent results were found for the $Countries$, $Impact$ and $Years$ variables and for the year in which the study was published.

3.5 Best practices

In the next step, this paper attempts to compute a mean estimate of the euro effect conditional on avoiding the medal mistakes described by Baldwin & Taglioni (2007) using an up-to-date methodology and control variables. Defining "best practices" is subjective since different studies may have different ideas about what best practices should be. However, the gravity equation is a widely researched area, and we have a great deal of empirical evidence at hand to help us define best practices. Thus, I believe that there is an added value of correcting the mean reported coefficients for the marginal effects of method choices.

For each variable presented in Table 4.1, a preferred value is selected (or

Figure 3.4: Model inclusion in Bayesian model averaging



Notes: These are the estimation results of equation 3.5 based on best 5000 models. Response variable: t-statistic of the euro coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one). All moderators are divided by the standard error of the euro coefficient estimate. Columns denote individual models, and variables are sorted by posterior inclusion probability in descending order from the top. Blue colour (darker in grayscale) = the variable is included and the estimated sign is positive. Red colour (lighter in grayscale) = the variable is included and the estimated sign is negative. White colour = 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 4.4. A detailed description of all variables is available in Table 4.1.

Table 3.6: Explaining the differences in the estimates of the euro effect

Response variable:	Bayesian model averaging			WLS			
T-statistic	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value	VIF
Data characteristics							
Midyear of data	0	0	0.623				
Disaggregated	-0.06	0.007	1	-0.019	0.019	0.335	21.08
Data IMF	-0.04	0.006	1	-0.051	0.018	0.005	12.61
Robustness check	0.005	0.006	0.419	0.001	0.018	0.945	5.25
Dependent variable	0.002	0.005	0.226	0.008	0.004	0.066	1.88
No. of years	0	0	0.036	-0.001	0.001	0.548	7
Years in EMU	0	0	0.026	0.001	0.005	0.797	95.17
Obs. per year	-0.006	0.001	1	-0.013	0.004	0.003	108.8
Crisis	-0.038	0.007	0.995	-0.064	0.034	0.062	31.56
EU data only	-0.001	0.002	0.068	-0.021	0.018	0.223	4.19
Total trade	-0.006	0.009	0.367	0.004	0.016	0.808	10.16
Countries	0	0	0.815	0	0	0.918	6.41
Treatment of multilateral resistance							
PPML	0	0.003	0.028	0.011	0.026	0.656	1.57
Time-varying fixed eff.	0.089	0.005	1	0.09	0.036	0.012	16.62
No control for MR	0.001	0.005	0.091	0.003	0.017	0.84	9.16
Anderson and Wincoop	0.001	0.003	0.101	0.026	0.015	0.071	20.07
Control variables							
FTA control	-0.078	0.008	1	-0.079	0.024	0.001	72.9
EU dummy	0	0.002	0.04	-0.012	0.022	0.586	29.28
Missing control	-0.011	0.012	0.522	0.008	0.027	0.753	13.19
Real exchange rate	-0.009	0.008	0.639	-0.02	0.01	0.044	7.99
Volatility	0	0.001	0.021	0.03	0.017	0.071	8.81
Publication characteristics							
Working paper	-0.028	0.006	1	0.028	0.022	0.216	30.36
Impact	-0.034	0.011	0.959	-0.007	0.035	0.835	23.21
Citations	0.029	0.006	1	0.04	0.02	0.049	141.5
Publication year	-0.124	0.017	1	0.207	0.106	0.055	67.13
Publication bias	0.008	0.325	0.423	-9.801	11.674	0.401	197.26
Havránek	-0.038	0.011	0.966	-0.001	0.034	0.967	95.27
Constant	250.655	NA	1	1.705	0.305	0	NA
Studies	57			57			
Observations	3,323			3,323			

Notes: Response variable: t-statistic of the euro coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for). All moderators are divided by the standard error of the euro coefficient estimate except for publication year, which is interacted with SE in the original model. PIP = posterior inclusion probability. SD = standard deviation. VIF = variance inflation factor. A detailed description of all variables is available in 4.1.

the value is left unchanged for a given estimate if this study does not have a preference for the value of the variable), and the implied semi-elasticity is computed as the mean predicted estimate of semi-elasticity. Essentially, this approach creates a synthetic study with the best practice methodology using a very large number of observations and the most recent data, following previous meta-analyses such as Havranek & Irsova (2011). This study selects sample maxima for mid-year data (more recent data), disaggregated data, the number of observations per year, whether the study was published, the number of citations, the impact factor, whether the study used volatility and exchange rates, time-varying fixed effects, and whether the study used FTA and the EU dummy. Sample minima are used for the dummy variables representing the type of value (in the main text), the dependent variable, missing control variables, total trade (summing trade flows before taking logs), and a lack of control for multilateral resistance. For all other variables, the actual values of the sample are kept.

The results are presented in Table 4.5. The overall mean semi-elasticity is reported in the last row. The column labelled "Difference from mean" shows the difference between the calculated predictions and the simple means that are reported in Table 3.1. The results are very similar – the overall mean semi-elasticity is approximately 0.03 but is not significantly different from zero. The precision of our best-practice estimate reflects the uncertainty of estimates of regression parameters of the WLS model. However, there might be even higher uncertainty connected with definition of the best-practice values of several variables, which would make the confidence intervals even wider.

Table 3.7: Best practice predictions are not significant from zero

Sample	Estimate	95% conf. int.		Difference from mean
Older studies (WLS)	0.031	-0.081	0.143	0.111
New studies (WLS)	0.027	-0.276	0.330	0.037
Full sample (WLS)	0.061	-0.072	0.195	0.039
Full sample (BMA)	0.027	-0.107	0.161	0.073

Notes: The table prediction of estimates of the euro coefficient for selected groups implied by BMA and WLS and our definition of best practice. That is, we take the regression coefficients estimated by BMA (Table 4.4) and predict the values of *euro* conditional on aspects of methods and data. Difference from mean = the difference between these estimates and the simple means reported in Table 3.1. The confidence intervals are approximate and constructed using the standard errors estimated by WLS.

3.6 Conclusion

The international economics literature focusing on the effects of currency unions, specifically on the effect of the euro on bilateral trade, has evolved substantially over the past decade, and this meta-analysis provides new insights on the factors that influence research outcomes. First, I show that publication bias has diminished over time, which is a positive result that contradicts a previous meta-analytic study by Havránek (2010). In addition, I find that the euro has a positive effect on trade and that this result is supported by almost 3500 estimates from over 50 studies. The number of estimates makes this study the largest meta-analysis to date. Compared to previous studies, I shifted the research focus from authorship to data and methodology, making my conclusions more rigorous. I expected several differences between studies that do not account for multilateral resistance and others based on the arguments of Baldwin & Taglioni (2007), who labels these issues as gold, silver, and bronze medals, but the meta-regression analysis based on the BMA exercise did not provide such evidence. On the other hand, my results show the influence of data sources, data characteristics, control variables, and methods rather than problems with multilateral resistance. I estimate the euro effect on bilateral trade to be between 2 and 6%, which is much smaller than that reported in the most recent study by Rose (2016). If we took just the new evidence (studies published after Havránek (2010)) or used best-practice predictions, we would not find any effect of euro on bilateral trade at all. This effect might be driven by the inclusion of the GFC period in the dataset and ongoing integration within the EU that makes it impossible to distinguish the effect of common currency from other effects in the longer time horizon since euro adoption.

Bibliography

Alakbarov, N. (2012): "Estimating the effect of common currencies on trade: The case study of eu and turkey." *Technical Report 60464*.

Alinaghi, N. & W. R. Reed (2018): "Meta-analysis and publication bias: How well does the fat-pet-peese procedure work?" *Research synthesis methods* **9(2)**: pp. 285–311.

Anderson, J. E. & E. Van Wincoop (2003): "Gravity with gravitas: a solution to the border puzzle." *the 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.

Aristotelous, K. (2006): "Are there differences across countries regarding the effect of currency unions on trade? evidence from emu." *Journal of Common Market Studies* **44(1)**: pp. 17–27.

Badinger, H. & K. Türkcan (2014): "Currency unions, export margins, and product differentiation: An empirical assessment for european monetary union." *Review of International Economics* **22(1)**: pp. 13–30.

Baldwin, R. (2006): "The euro's trade effect." *Working Paper Series 594*, European Central Bank.

Baldwin, R., V. DiNino, L. Fontagné, R. A. D. Santis, & D. Taglioni (2008): "Study on the impact of the euro on trade and foreign direct investment." *European Economy - Economic Papers 321*, Directorate General Economic and Monetary Affairs, European Commission.

Baldwin, R., F. Skudelny, & D. Taglioni (2005): "Trade effects of the euro - evidence from sectoral data." *Working Paper Series 446*, European Central Bank.

Baldwin, R. & D. Taglioni (2004): "Positive oca criteria: Microfundations for the rose effect." *COE/RES Discussion Paper Series 34*, COE/RES.

Baldwin, R. & D. Taglioni (2007): "Trade effects of the euro: A comparison of estimators." *Journal of Economic Integration* **22(4)**: pp. 780–818.

Barr, D., F. Breedon, & D. Miles (2003): "Life on the outside: economic conditions and prospects outside euroland." *Economic Policy* **18(37)**: pp. 573–613.

Berger, H. & V. Nitsch (2008): "Zooming out: The trade effect of the euro in historical perspective." *Journal of International Money and Finance* **27(8)**: pp. 1244–1260.

Bergin, P. R. & C.-Y. Lin (2012): "The dynamic effects of a currency union on trade." *Journal of International Economics* **87(2)**: pp. 191–204.

Brouwer, J., R. Paap, & J.-M. Viaene (2008): "The trade and fdi effects of emu enlargement." *Journal of International Money and Finance* **27(2)**: pp. 188–208.

Bun, M. J. G. & F. J. G. M. Klaassen (2002): "Has the euro increased trade?" *Tinbergen Institute Discussion Papers 02-108/2*, Tinbergen Institute.

Bun, M. J. G. & F. J. G. M. Klaassen (2007): "The euro effect on trade is not as large as commonly thought." *Oxford Bulletin of Economics and Statistics* **69(4)**: pp. 473–496.

Buongiorno, J. (2015): "Monetary union and forest products trade - the case of the euro." *Journal of Forest Economics* **21(4)**: pp. 238–249.

Cafiso, G. (2011): "Rose effect versus border effect: the euro's impact on trade." *Applied Economics* **43(13)**: pp. 1691–1702.

Camarero, M., E. Gómez, & C. Tamarit (2014): "Is the 'euro effect' on trade so small after all? new evidence using gravity equations with panel cointegration techniques." *Economics Letters* **124(1)**: pp. 140–142.

Campbell, D. L. & A. Chentsov (2017): "Breaking badly: The currency union effect on trade."

Chintrakarn, P. (2008): "Estimating the euro effects on trade with propensity score matching." *Review of International Economics* **16(1)**: pp. 186–198.

Cieślik, A., J. J. Michałek, & J. Mycielski (2012): "Euro and trade flows in central europe." *Equilibrium* **7(3)**: pp. 7–25.

Cieślik, A., J. J. Michałek, & J. Mycielski (2014): "Trade effects of the euro adoption by the eu new member states."

Cipollina, M. & L. Salvatici (2010): "Reciprocal trade agreements in gravity models: A meta-analysis." *Review of International Economics* **18(1)**: pp. 63–80.

Costa-Font, J. (2010): "Regional single currency effects on bilateral trade with the european union." *LEQS Paper* **(26)**: pp. –.

Davidson, R. & J. G. MacKinnon (2004): *Econometric theory and methods*, volume 5. Oxford University Press New York.

Doucouliagos, C. & 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. Henn (2011): "One money, one market: A revised benchmark." *Review of International Economics* **19(3)**: pp. 419–435.

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.

Faruqee, H. (2004): "Measuring the trade effects of emu." *IMF Working Papers 04/154*, International Monetary Fund.

Feldkircher, M. & S. Zeugner (2009): "Benchmark priors revisited: on adaptive shrinkage and the supermodel effect in bayesian model averaging." *IMF Working Papers* **(9-202)**.

Fernandes, A. (2006): "Trade dynamics and the euro effect: Sector and country estimates." University of Essex and IPVC.

Flam, H. & H. Nordström (2006a): "Euro effects on the intensive and extensive margins of trade." *CESifo Working Paper Series 1881*, CESifo Group Munich.

Flam, H. & H. Nordström (2006b): "Trade volume effects of the euro: Aggregate and sector estimates." *Seminar Papers 746*, Stockholm University, Institute for International Economic Studies.

Flam, H. & H. Nordström (2007): "The euro and single market impact on trade and fdi." Mimeo.

Flam, H. & H. Nordström (2008): "The euro impact on fdi revisited and revised." Mimeo.

Frankel, J. (2010): "The estimated trade effects of the euro: Why are they below those from historical monetary unions among smaller countries?" In "Europe and the Euro," NBER Chapters, pp. 169–212. National Bureau of Economic Research, Inc.

Glick, R. & A. K. Rose (2015): "Currency unions and trade: a post-emu mea culpa." *Technical report*, National Bureau of Economic Research.

Glick, R. & A. K. Rose (2016): "Currency unions and trade: A post-emu reassessment." *European Economic Review* .

Gomes, T., J. Helliwell, T. Kano, & L. S. J. Murra and (2006): "Bank of canada working paper." *Wp*, Bank of Canada Working Paper.

Havranek, T. (2015): "Measuring intertemporal substitution: The importance of method choices and selective reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

Havranek, T., D. Herman, & Z. Irsova (2018a): "Does daylight saving save energy? a meta-analysis." *Energy Journal* **39(2)**: pp. 35–61.

Havranek, T., R. Horvath, Z. Irsova, & M. Rusnak (2015): "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

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, & T. Vlach (2018b): "Measuring the income elasticity of water demand: The importance of publication and endogeneity biases." *Land Economics* **forthcoming**.

Havránek, T. (2010): "Rose effect and the euro: Is the magic gone?" *Review of World Economics* **146(2)**: pp. 241–261.

Havránek, T. & Z. Iršová (2017): "Do borders really slash trade? a meta analysis." *IMF Economic Review* **65(2)**: pp. 365–396.

Havránek, T., Z. Iršová, & K. Janda (2012): "Demand for gasoline is more price-inelastic than commonly thought." *Energy Economics* **34(1)**: pp. 201–207.

Havránek, T., M. Rusnák, & A. Sokolova (2017): "Habit formation in consumption: A meta-analysis." *European Economic Review* **95**: pp. 142–167.

Head, K. & T. Mayer (2014): "Gravity equations: Workhorse, toolkit, and cookbook." *Handbook of International Economics* **4**.

Iršová, Z. & T. Havránek (2013): "Determinants of horizontal spillovers from fdi: Evidence from a large meta-analysis." *World Development* **42**: pp. 1–15.

Jung, B., J. Hogrefe, & W. Kohler (2010): "Readdressing the trade effect of the euro: Allowing for currency misalignment." *ZEW Discussion Papers 10-023*, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Kelejian, H., G. S. Tavlas, & P. Petroulas (2012): "In the neighborhood: The trade effects of the euro in a spatial framework." *Regional Science and Urban Economics* **42(1)**: pp. 314–322.

Kunroo, M. H., I. A. Sofi, & N. A. Azad (2016): "Trade implications of the euro in emu countries: a panel gravity analysis." *Empirica* **43(2)**: pp. 391–413.

Kunrooa, M. H. & N. A. Azad (2015): "Impact of single currency adoption on trade flows: The panel data evidence." *International Journal of Economics and Empirical Research (IJEER)* **3(9)**: pp. 433–439.

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

Martínez-Zarzoso, I. & F. Johannsen (2017): "Euro effect on trade in final, intermediate and capital goods." *International Journal of Finance & Economics* **22(1)**: pp. 30–43.

McGowan, D. (2008): "Has the euro increased trade?" *Michigan journal of business* **1(2)**: pp. 11–52.

Micco, A., E. Stein, & G. Ordoñez (2003): "The currency union effect on trade: early evidence from emu." *Economic Policy* **18(37)**: pp. 315–356.

Mika, A. & R. Zymek (2017): "Friends without benefits? new emu members and the 'euro effect'on trade." .

Moreno, S. G., A. J. Sutton, A. Ades, T. D. Stanley, K. R. Abrams, J. L. Peters, & N. J. Cooper (2009): "Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study." *BMC medical research methodology* **9(1)**: p. 1.

Murphy, G. & I. Siedschlag (2011): "Has the euro boosted intra-euro area exports? evidence from industry data." *Review of economics and institutions* **2(3)**: pp. 22–.

de Nardis, S., R. de Santis, & C. Vicarelli (2008a): "The euro's effects on trade in a dynamic setting." *European Journal of Comparative Economics* **5(1)**: pp. 73–85.

de Nardis, S. & C. Vicarelli (2003a): "Currency unions and trade: The special case of emu." *Review of World Economics* **139**: pp. 625–649.

de Nardis, S. & C. Vicarelli (2003b): "The impact of euro on trade: The (early) effect is not so large." *ISAE Working Papers 31*, ISAE - Institute for Studies and Economic Analyses - (Rome, ITALY).

de Nardis, S., C. Vicarelli, & R. De Santis (2008b): "The single currency's effects on eurozone sectoral trade: Winners and losers?" *Economics - The Open-Access, Open-Assessment E-Journal* **2(17)**: pp. 1–34.

Nähle, T. (2015): "Exchange rates and trade: The impacts of the euro on bilateral export flows."

Polák, P. (2016): "The productivity paradox: A meta-analysis." *Information Economics and Policy* **forthcomming**: pp. –.

Polyak, O. (2014): "The impact of euro adoption on export performance: Comparison of the czech republic and slovakia." *Technical report*.

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. (2008): "Is emu becoming an optimum currency area? the evidence on trade and business cycle synchronization." *Technical report*, University of California, Berkeley.

Rose, A. K. (2016): "Why do estimates of the emu effect on trade vary so much?" *Open Economies Review* pp. 1–18.

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.

Rotili, L. (2014): "The euro effects on intermediate and final exports." *Technical Report 7/14*, Sapienza University of Rome WP.

Ruiz, J. M. & J. M. Vilarrubia (2007): "The wise use of dummies in gravity models: export potentials in the euromed region." *Banco de España Research Paper No. WP-0720* pp. –.

Serlenga, L. & Y. Shin (2013): "The euro effect on intra-eu trade: Evidence from the cross sectionally dependent panel gravity models." In "Fifth Italian Congress of Econometrics and Empirical Economics," pp. –.

Shin, Y. & L. Serlenga (2007): "Gravity models of intra-eu trade: application of the ccep-ht estimation in heterogeneous panels with unobserved common time-specific factors." *Journal of Applied Econometrics* **22(2)**: pp. 361–381.

Silva, J. M. C. S. & S. Tenreyro (2010): "Currency unions in prospect and retrospect." *CEP Discussion Papers dp0986*, Centre for Economic Performance, LSE.

de Sousa, J. & J. Lochard (2009): "The currency union effect on trade and the fdi channel." *Cahiers de la Maison des Sciences Economiques j04111*, Université Panthéon-Sorbonne (Paris 1).

de Souza, L. V. (2002): "Trade effects of monetary integration in large, mature economies: A primer on the european monetary union." *Kiel Working Papers 1137*, Kiel Institute for the World Economy.

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. (2017): "Limitations of pet-peese and other meta-analysis methods." *Social Psychological and Personality Science* **8(5)**: pp. 581–591.

Stanley, T. D. & C. 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., C. 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. & H. Doucouliagos (2019): "Practical significance, meta-analysis and the credibility of economics." .

Stanley, T. D., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh *et al.* (2013): "Meta-analysis of economics research reporting guidelines." *Journal of Economic Surveys* **27(2)**: pp. 390–394.

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.

Taglioni, D. (2004): *Monetary union, exchange rate variability and trade, dissertation thesis*. Master's thesis, Université de Genéve.

Vicarelli, C. & C. Pappalardo (2012): "Euro introduction and export behaviour of italian firms." *ISTAT, Italian National Institute of Statistics* .

Viscusi, W. K. (2017): "Best estimate selection bias of estimates of the value of a statistical life." *Technical report*, Vanderbilt Law Research Paper No. 17-7.

Westphal, H. (2013): *Are There Winners and Losers in the Game of the Euro?* Master's thesis, Lund University.

Zigraiova, D. & T. Havranek (2016): "Bank competition and financial stability: Much ado about nothing?" *Journal of Economic Surveys* **30(5)**: pp. 944–981.

3.A Appendix

This appendix provides additional statistical tests of publication bias and robustness check of explanatory analysis of the heterogeneity of results in the collected estimates of the euro effects. This section also provides the list of studies included in the meta-analysis together with the indication if that paper was used by Havránek (2010) or not.

3.A.1 Detailed funnel plots

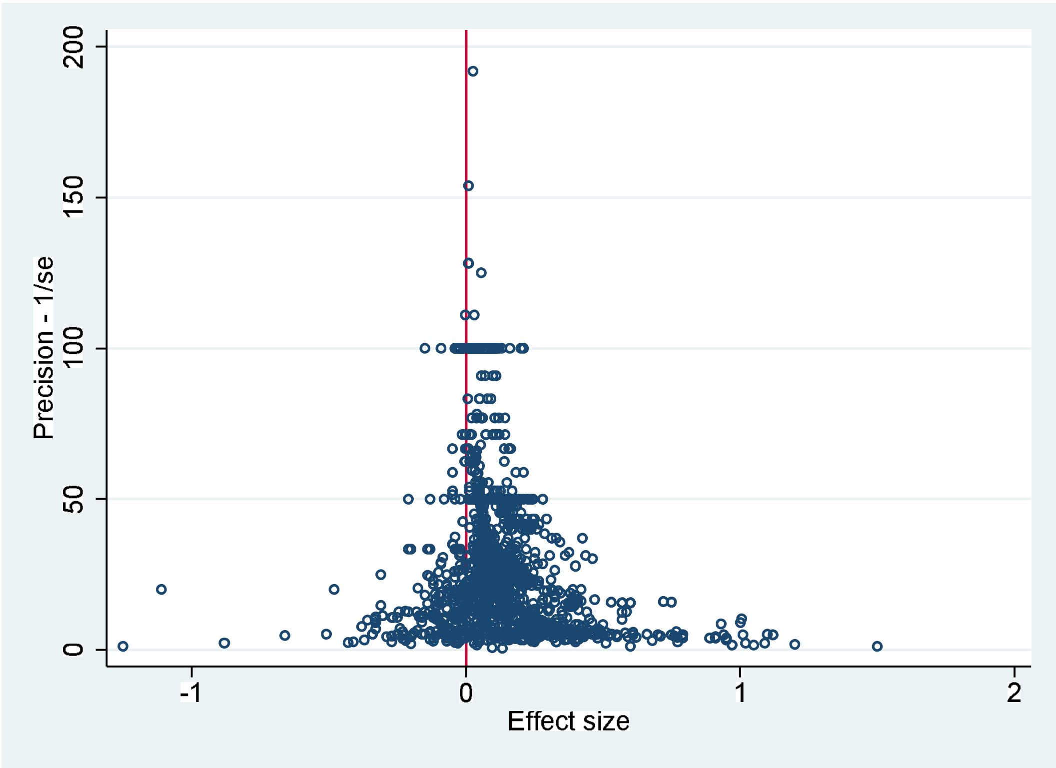


Figure 3.5: Funnel plot – only studies examined by Havránek (2010)

Notes: The funnel plot is a scatter plot where estimate (Effect size) is plotted against its precision (inverse of standard error). In the absence of publication bias the funnel should be symmetrical around the value given by estimates with the highest precision. The vertical line denotes the zero value, mean and median values are reported in Table 3.1.

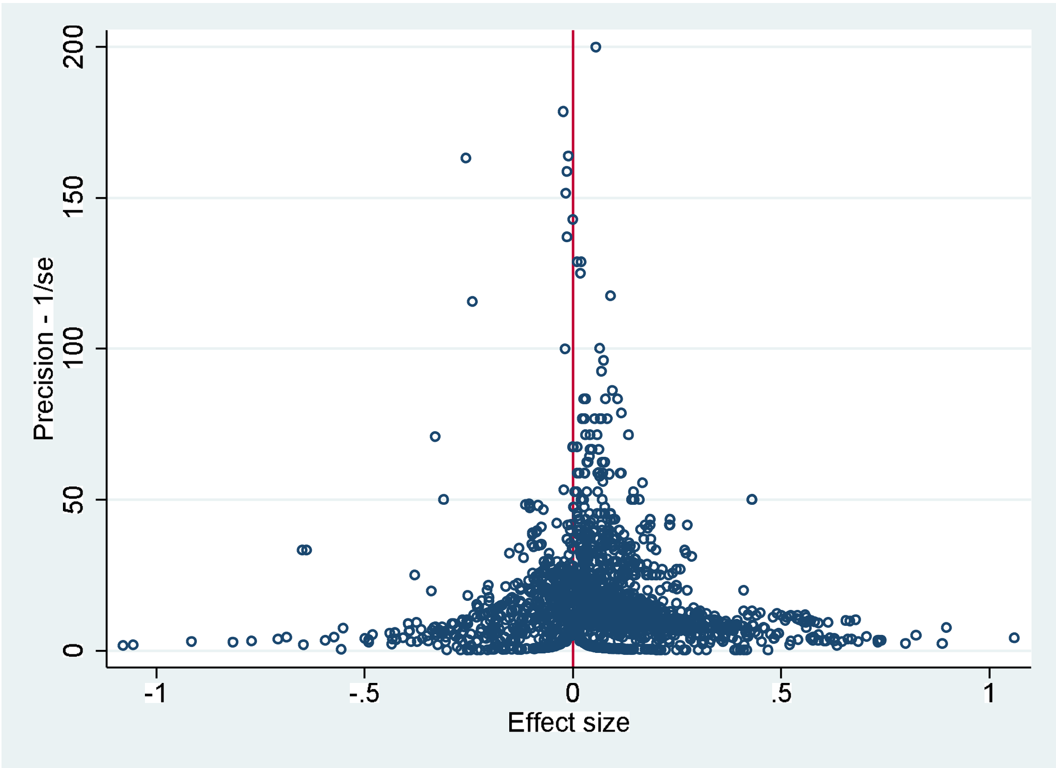


Figure 3.6: Funnel plot – only studies not examined by Havránek (2010)

Notes: The funnel plot is a scatter plot where estimate (Effect size) is plotted against its precision (inverse of standard error). In the absence of publication bias the funnel should be symmetrical around the value given by estimates with the highest precision. The vertical line denotes the zero value, mean and median values are reported in Table 3.1.

3.A.2 FAT-PET-PEESE

Robustness check with CU (currency union) units. The following FAT-PET-PEESE tests are performed using calculated currency union effects, which are calculated as $CU = e^{\gamma} - 1$, with γ being the estimate of semi-elasticity of the euro effect from the gravity equation. Therefore, CU is the value of the estimated euro effect in percentages. The results of presented in Tables 3.8, 3.9 and 3.10 and consistent with results presented in Tables 3.3, 3.5, 3.4. This supports the robustness of the presented findings about presence of publication bias and its decreasing trends over time.

Table 3.8: FAT-PET-PEESE with CU for whole dataset

	FAT-PET	PEESE	IV	FE
Precision – 1/se	0.0281***	0.0510***	0.0324***	0.00187
(Effect beyond bias)	(2.96)	(30.17)	(3.40)	(0.18)
Standard error		0.238**		
(Publication bias)		(2.50)		
Constant	1.133***		1.087***	1.710***
(Publication bias)	(2.81)		(2.85)	(7.64)
Observations	3322	3322	3322	3322

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; t statistics in parentheses; the table presents the results of regressions 3.4 and 3.6; dependent variable: t-statistic of CU = exp(euro coefficient)-1; standard errors clustered by study; FE = fixed effects; IV = instrumental variable, instrument is the inverse of logarithm of the number of observations per year.

Table 3.9: FAT-PET-PEESE with CU for new studies

	FAT-PET	PEESE	IV	FE
Precision – 1/se	0.00213	0.0244***	0.0130	-0.0220
(Effect beyond bias)	(0.11)	(8.71)	(0.86)	(-1.12)
Standard error		0.101		
(Publication bias)	(1.16)			
Constant	0.844**		0.729**	1.233***
(Publication bias) (2.18)		(2.30)	(3.89)	
Observations	1741	1741	1741	1741

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; t statistics in parentheses; the table presents the results of regressions 3.4 and 3.6; dependent variable: t-statistic of CU = exp(euro coefficient)-1; standard errors clustered by study; FE = fixed effects; IV = instrumental variable, instrument is the inverse of logarithm of the number of observations per year.

Table 3.10: FAT-PET-PEESE with CU for older studies

	FAT-PET	PEESE	IV	FE
Precision – 1/se	0.0283***	0.0598***	0.0327***	0.0158
(Effect beyond bias)	(4.10)	(28.00)	(3.68)	(1.52)
Standard error		4.170***		
(Publication bias)		(8.33)		
Constant	1.902***		1.797***	2.260***
(Publication bias)	(3.78)		(3.54)	(7.56)
Observations	1581	1581	1581	1581

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; t statistics in parentheses; the table presents the results of regressions 3.4 and 3.6; dependent variable: t-statistic of CU = exp(euro coefficient)-1; standard errors clustered by study; FE = fixed effects; IV = instrumental variable, instrument is the inverse of logarithm of the number of observations per year.

3.A.3 BMA Robustness check

This section provides two additional sets of results, which are estimates as a robustness check of the presented BMA exercise. First, I use alternative priors for the Bayesian model averaging. Second, we employ weighted least squares (WLS) regression using only variables with Posterior inclusion probability (PIP) higher than 0.3. This exercise shoes that the results are similiar to the main results – in terms of estimated coefficients for the explanatory variables.

The baseline specification presented in Section 3.4.2 the paper employs intuitive priors, which according to Eicher *et al.* (2011) do yield the best predictive performace. The first is the uniform model prior, which gives each model the same prior probability. Any model is therefore not thought to be more probable than the others ex-ante. Second choice is about the information prior. In the baseline specification again the unit information prior for Zellner's g-prior is used. It means that this prior (each regression coefficient equals zero) provides the same abour of information as one observation in the data. Because we have more than 3,300 observations, the prior does not drive the posterior results. There are many other possible priors and their choices might influence the results.

The disadvantage of the uniform model prior is that ti gives more weight to models with the mean number of variables, which is 14 in our case. But the true model may include only a few variables, which makes the uniform approach less productive. Alternative prior would give the same prior probability to each model size, which means that there will be no preference about model size (Ley & Steel, 2009). Frequently used alternative information prior is BRIC g-prior (Havránek & Iršová, 2017). This study uses these alternatives for the robustness check which results are presented in Table 3.11.

Table 3.11: Robustness check - alternative priors for BMA

Response variable:	Bayesian model averaging			WLS		
T-statistic	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value
Data characteristics						
Midyear of data	0	0	0.257			
Disaggregated	-0.046	0.007	1	-0.024	0.016	0.127
Data IMF	-0.034	0.007	1	-0.04	0.02	0.051
Robustness check	0.006	0.006	0.508	0.005	0.019	0.807
Dependent variable	0.003	0.006	0.315	0.007	0.004	0.086
No. of years	0	0	0.022			
Obs. per year	-0.008	0.001	1	-0.01	0.004	0.022
Crisis	-0.042	0.008	1	-0.04	0.015	0.009
EU data only	0	0.002	0.029			
Total trade	-0.001	0.003	0.062			
Countries	0	0	0.981	0	0	0.882
Treatment of multilateral resistance						
PPML	0	0.002	0.012			
Time-varying fixed eff.	0.088	0.005	1	0.091	0.029	0.002
No control for MR	0.004	0.009	0.171			
Anderson and Wincoop	0	0.002	0.06			
Control variables						
FTA control	-0.068	0.007	1	-0.072	0.013	0
EU dummy	0	0.001	0.014			
Missing control	-0.006	0.011	0.256			
Real exchange rate	-0.012	0.008	0.752	-0.016	0.012	0.158
Volatility	0.012	0.009	0.694	0.026	0.013	0.05
Publication characteristics						
Working paper	-0.003	0.005	0.238			
Impact	0	0.002	0.021			
Citations	0.017	0.004	1	0.014	0.012	0.245
Publication year	0	0.001	0.226			
Publication bias	-0.349	1.354	0.239	0.132	0.025	0
Havránek	-0.003	0.007	0.194			
Constant	1	NA	1	1.735	0.306	0
Studies	61			61		
Observations	3,323			3,323		

Notes: Response variable: t-statistic of the euro coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for). All moderators are divided by the standard error of the euro coefficient estimate, except of publication year, which is interacted with SE in the original model. PIP = posterior inclusion probability. SD = standard deviation. In this specification we used beta-binomial prior presented by Ley & Steel (2009). In the WLS check we only include explanatory variables with PIP > 0.3. A detailed description of all variables is available in 4.1.

3.A.4 Studies included in the meta-analysis

Table 3.12: List of studies

Study	Published	Havránek	WP
Bun & Klaassen (2002)	2002	Yes	Yes
de Souza (2002)	2002	Yes	Yes
Barr *et al.* (2003)	2003	Yes	No
Micco *et al.* (2003)	2003	Yes	No
de Nardis & Vicarelli (2003b)	2003	Yes	Yes
de Nardis & Vicarelli (2003a)	2003	Yes	No
Baldwin & Taglioni (2004)	2004	Yes	Yes
Faruqee (2004)	2004	Yes	Yes
Taglioni (2004)	2004	Yes	Yes
Baldwin *et al.* (2005)	2005	No	Yes
Aristotelous (2006)	2006	Yes	No
Baldwin (2006)	2006	Yes	Yes
Fernandes (2006)	2006	No	Yes
Flam & Nordström (2006a)	2006	Yes	Yes
Flam & Nordström (2006b)	2006	Yes	Yes
Gomes *et al.* (2006)	2006	Yes	No
Baldwin & Taglioni (2007)	2007	Yes	No
Bun & Klaassen (2007)	2007	Yes	No
Flam & Nordström (2007)	2007	No	Yes
Ruiz & Vilarrubia (2007)	2007	No	Yes
Shin & Serlenga (2007)	2007	Yes	No
Baldwin *et al.* (2008)	2008	Yes	Yes
Berger & Nitsch (2008)	2008	Yes	No
Brouwer *et al.* (2008)	2008	Yes	No
Flam & Nordström (2008)	2008	No	Yes
Chintrakarn (2008)	2008	Yes	No
McGowan (2008)	2008	No	Yes
de Nardis *et al.* (2008b)	2008	Yes	No
de Nardis *et al.* (2008a)	2008	No	No
de Sousa & Lochard (2009)	2009	No	Yes
Costa-Font (2010)	2010	No	Yes
Frankel (2010)	2010	Yes	Yes
Jung *et al.* (2010)	2010	No	Yes
Silva & Tenreyro (2010)	2010	No	Yes
Cafiso (2011)	2011	Yes	No
Eicher & Henn (2011)	2011	No	No
Murphy & Siedschlag (2011)	2011	No	No
Alakbarov (2012)	2012	No	Yes

Continued on next page

Table 3.12: List of studies (continued)

Study	Published	Havránek	WP
Bergin & Lin (2012)	2012	No	No
Cieślik *et al.* (2012)	2012	No	No
Kelejian *et al.* (2012)	2012	No	No
Vicarelli & Pappalardo (2012)	2012	No	Yes
Serlenga & Shin (2013)	2013	No	Yes
Westphal (2013)	2013	No	Yes
Badinger & Türkcan (2014)	2014	No	No
Camarero *et al.* (2014)	2014	No	No
Cieślik *et al.* (2014)	2014	No	No
Polyak (2014)	2014	No	Yes
Rotili (2014)	2014	No	Yes
Buongiorno (2015)	2015	No	No
Kunrooa & Azad (2015)	2015	No	No
Nähle (2015)	2015	No	Yes
Glick & Rose (2016)	2016	No	No
Kunroo *et al.* (2016)	2016	No	No
Martínez-Zarzoso & Johannsen (2017)	2016	No	No
Campbell & Chentsov (2017)	2017	No	Yes
Mika & Zymek (2017)	2017	No	Yes

Notes: WP = working paper, Havránek = study included in Havránek (2010)

Chapter 4

How bad are trade wars? Evidence from trade costs

Abstract[1]

We use more than 1600 estimates from 71 studies to investigate the relation between international trade flows and trade costs, which are mainly tariffs. Our results suggest that the empirical literature suffers from the presence of publication bias, which has exaggerated the effect (the elasticity is higher). After accounting for publication bias, we estimate trade elasticity with respect to trade cost to be between -0.9 and -2.0. We also identified several properties that explain the heterogeneity of the reported estimates. The results from Bayesian model averaging, which takes into account model uncertainty, show that the differences among estimates are systematically driven by the data source, control variables, and estimation techniques. The mean reported estimate of the trade cost conditional on the best-practice approach is not statistically different from zero due to the large confidence interval.

[1]This paper is co-authored with Nikol Poláková and Anna Tlustá. The views expressed in this paper are those of the author and not necessarily those of the Czech National Bank.

4.1 Introduction

Trade is a keystone of economics. Trade allows us to exchange things that we have for things that we want. We have seen periods of protectionism in history, but the lesson learned is that any limitation of trade is not beneficial for any of the parties. On the contrary, abolition of trade barriers leads to an increase in both trade and welfare. Therefore, countries have entered into trade agreements and multinational World Trade Organization (WTO) deals with global rules of trade between nations with the main function being to ensure that trade flows as smoothly, predictably and freely as possible. Yet, the current developments in the world are leading in a different direction. Since February 2020, the United Kingdom is no longer part of the European Union (EU), and since mid 2018, the trade war between China and the USA – two of the world's largest economies – has been ongoing. The USA did not impose tariffs on only China but also on other trading partners. In reaction, China also imposed tariffs on US imports. The extent of the trade cost increased caused by the currently materialized increased trade tensions is unclear. For example, Stephens (2015) warns that there is a risk of returning to a beggar-thy-neighbor trade policy as we saw in 1930s, which resulted in recession.

International trade has been investigated by economists for decades. Economists have been searching for a variety of variables that have a significant effect on international trade. Growing literature about international trade provides evidence that trade costs have a negative impact on trade volume. Such findings confirm historical studies investigated by Anderson & Van Wincoop (2004), who show that integration leads to trade cost reduction. According to Estevadeordal *et al.* (2003) and De Bromhead *et al.* (2019), higher transport costs were the main determinant in the collapse of world trade in the 1930s.

Trade costs are not only related to tariffs and transport. We can group trade costs into several groups, including policy costs (currency, tariffs, security, etc.), distribution costs and transport costs. Currency is investigated not only from a currency area perspective (see, e.g., Havránek (2010) or Polák (2019)) but also from an exchange rate perspective; see, e.g., Ćorić & Pugh (2010). Non-financial focus is also crucial. Studies such as Baier & Bergstrand (2001) show that liberalization of international transport services promotes international trade to the same magnitude as tariff liberalization, which supports an up-to-date world trade strategy, where the role of governance and infrastructure is more important than the role of borders. An importer or exporter encounters trade costs during the whole process of trade, not only at the borders. The costs start with collecting information about market conditions and end with final payment. Bajzik *et al.* (2019) analyses the Armington elasticity or the relation between import and domestic prices as competitors. Minimization of trade costs through facilitation of merchandise and services trade logistics, both inbound and outbound, is currently the centre of attention.

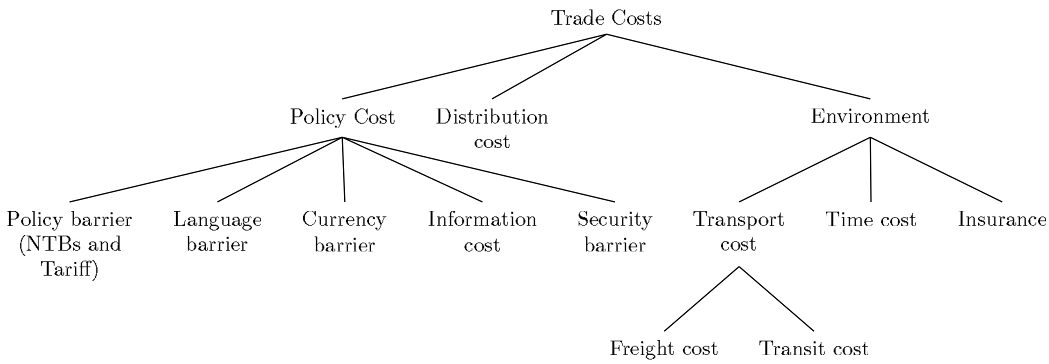
The most commonly used model to empirically estimate the effect of trade costs on international trade in academia is the gravity equation. The gravity framework has undergone significant development since it was first introduced in the second half of the 20th century. In this area, the key papers, such as Head & Mayer (2014) and Anderson & Van Wincoop (2003), point to the weaknesses of some approaches used and suggest remedies to overcome the issues. To empirically evaluate the existing literature, consisting of more than 70 papers, we apply a meta-analytic framework to investigate the effect of trade costs on bilateral trade flows, focusing on the methodology to determine whether different approaches result in different results in reality.

The structure of this paper is as follows: Section 4.2 discusses potential discrepancies among different kinds of trade costs. Section 4.3 describes how trade costs are investigated empirically using the gravity framework. Section 4.4 presents the data collected from the studies. Section 4.5 discusses publication bias, and Section 4.6 discusses our results – the key drivers of the heterogeneity of trade cost estimates. Finally, Section 4.7 concludes.

4.2 Trade costs importance

In general, we can define *trade costs* as all costs that are necessary to get a good to a final consumer, except the marginal cost of the production of the good itself. This includes transportation costs, policy barriers, legal and regulatory costs, and distribution costs. Trade costs are reported mostly in terms of their ad valorem tax equivalent. Anderson & Van Wincoop (2004) highlight the importance of trade cost with the statement that the *"Tax equivalent of representative trade costs for rich countries is 170 percent. This includes all transport, border-related and local distribution costs from foreign producer to final user in the domestic country. Trade costs are richly linked to economic policy. Direct policy instruments (tariffs, the tariff equivalents of quotas and trade barriers associated with the exchange rate system) are less important than other policies (transport infrastructure investment, law enforcement and related* *property rights institutions, informational institutions, regulation, language)"*. Figure 4.1 shows one possible division of trade costs.

Figure 4.1: The division of trade costs



Notes: The figure presents an illustrative division of trade costs, as can be found in the literature.

Policy costs or policy barriers to trade are intuitive; they include tariffs, quotas, license fees, exchange rates, and security issues. Anderson & Van Wincoop (2004) notes that even though it should be simple to determine how high the policy costs of trade are, they are difficult to examine empirically because of missing and incomplete data. Unavailability and poor quality of data with respect to the policy costs of trade is surprising since it has been known for hundreds of years that open trade promotes welfare. Researchers often use the United Nations Conference on Trade and Developments Trade Analysis & Information System (TRAINS) dataset since it is easily accessible, but there are other measures and indexes, as noted in, e.g., Hoekman & Nicita (2008).

Distribution costs are cost belonging to wholesale and retail distribution that affect the final prices of goods in each region or country. Burstein *et al.* (2003), in their study on distribution costs and real exchange rates, state that distribution costs are key drivers of different prices of the same good in different countries because distribution services are focused on labour and land costs. In addition, their study shows variation in distribution margins exists even within the same country. We might expect that increasing competition and possible arbitrage potential would cause the distribution cost to be the smallest part of the overall cost, but the opposite is true – the margins for distribution are usually the highest.

Environment costs are costs raised mainly by distance. Aside from time cost and insurance, the main component is transport costs. Direct transport costs include freight rates (the transit cost itself depends on the type of transport) and insurance. Indirect transport costs include the cost for transport preparation and the cost for having goods in transit. Transport costs are higher for heavier goods, like agriculture products, but according to Limao & Venables (2001), the measurement of transport costs is easier than policy costs, and transport cost are comparable across countries and commodities, on average.

Figure 4.1 and a brief description inspired by the extensive survey of Anderson & Van Wincoop (2004) provides a general picture, where the term "trade cost" is very general and can measure many of different factors. On the other hand, research is limited by the available data; therefore, final empirical research is quite homogeneous in terms of definitions and methodology.

4.3 Theoretical background

The gravity equation is widely used in international economics and is said to date back to the nineteenth century, when Ravenstein (1885) used a gravity-type relationship to study immigration. The model was formalized in the 1960s when Tinbergen (1962) and Pöyhönen (1963) applied it to study the impact of trade policy. In the early models, bilateral trade flows depend on only the GDP of the countries and the distance between them, so the relationship has the same interpretation as in physics:

$$T_{ij} = G\frac{Y_i^{\alpha} Y_j^{\beta}}{D_{ij}^{\theta}} \tag{4.1}$$

where T_{ij} is the trade flow from region or country i to destination j and Y_i and Y_j are the economic sizes of the regions or countries (mostly represented by GDP). D_{ij} is the distance between the regions or countries, G replaces the gravity constant and hence represents all other bilateral indicators, and coefficients α, β and θ stand for the elasticities to be estimated.

The first theoretical foundations of the gravity equation were provided by Anderson (1979) and Bergstrand (1985). At the beginning of the 21st century, the most influential works by Eaton & Kortum (2002) and Anderson & Van Wincoop (2003) gave the gravity framework a more solid foundation. Eaton & Kortum (2002) derived the gravity model from a Ricardian supply-side framework, while Anderson & Van Wincoop (2003) derived the gravity model from an Armington demand-side framework (CES production) and emphasized the importance of the general equilibrium effects of trade costs. Although these works approach the subject from different perspectives, the results are almost the same. The fact that many more frameworks can be used to derive the gravity equation was presented by Arkolakis *et al.* (2012).

We present the gravity equation based on the Armington framework following Anderson & Van Wincoop (2003) since that form is commonly used in the most recent studies to analyse the impact of different trade costs on international trade. The complete generalized gravity system that explains exports (X_{ij}) from country i to j is described according to Yotov *et al.* (2016) in equations 4.2 to 4.4:

$$X_{ij} = Y_i E_j \left(\frac{\tau_{ij}}{\Pi_i P_j} \right)^{1-\omega} \tag{4.2}$$

$$\Pi_i = \left(\sum_{j=1}^{C} E_j \left(\frac{\tau_{ij}}{P_j} \right)^{1-\omega} \right)^{\frac{1}{1-\omega}} \tag{4.3}$$

$$P_j = \left(\sum_{i=1}^{C} Y_i \left(\frac{\tau_{ij}}{\Pi_i} \right)^{1-\omega} \right)^{\frac{1}{1-\omega}} \tag{4.4}$$

where ω is the broadly defined *trade cost elasticity* and τ_{ij} is the total bilateral trade cost.

Equation 4.2 is the structural gravity equation that governs bilateral trade flows and consists of a size term $Y_i E_j$ and a trade cost term $\frac{\tau_{ij}}{\Pi_i P_j}$. The size term is measured as the exporter and importer incomes. The innovation brought by Anderson & Van Wincoop (2003) is the multilateral resistance terms P_j and Π_i that differentiate the theory-founded gravity models from earlier versions. These remoteness terms represent the ease of market access for importers and exporters.

To provide econometric estimates for the effect of distance and other bilateral indicators and measures of trade costs, trade economists use the multiplicative nature of gravity equation and apply a log form of Equation 4.1. Because of the remarkable predictive power of the gravity model and its intuitive form, most of the gravity-related literature is empirical. Moreover, as stated above, the empirical applications were predecessors of the theoretical foundations. The log-linear form of 4.2 for any time t can be rewritten as follows:

$$\ln X_{ij,t} = \ln E_j, t + \ln Y_i, t + (1-\omega)\ln \tau_{ij,t} - (1-\omega)\ln P_{j,t} - (1-\omega)\ln \Pi_{i,t} + \epsilon_{ij,t} \tag{4.5}$$

where the error term is ϵ_{ij}. The specification 4.5 is the most frequently used and estimated version of the empirical gravity equation. Yet, some obvious challenges occur when estimating this equation. First, the multilateral resistance terms P_j and Π_i are theoretical and are not directly observable. Baldwin & Taglioni (2007) emphasise the importance of proper control for the multilateral resistance terms; failure to control the terms is called the "Gold Medal Mistake". One remedy is the application of directional (exporter and importer) fixed effects; additionally, when using panel data, exporter-time and importer-time fixed effects are necessary. However, fixed effects absorb all observable and unobservable country-specific characteristics.

We are interested in trade costs, the decomposition of which represents another challenge. The bilateral trade costs term τ_{ij} is disaggregated using a group of observable variables that have become standard in the empirical gravity equation:

$$\begin{aligned}(1-\omega)\ln \tau_{ij,t} = {} & \beta_1 \ln Distance_{ij} + \beta_2 Language_{ij} + \beta_3 Border_{ij} + \beta_4 Currency_{ij,t} + \\ & + \beta_5 RTA_{ij,t} + \beta_6 \ln(1 + tariff_{ij,t})\end{aligned} \tag{4.6}$$

where first three variables capture time-invariant properties: $Distance$ is the distance between trading countries, $Language$ is a dummy variable for a common official language and $Border$ is a dummy variable for a common border. Then, we have $Currency$, which is a dummy for the same currency at time t and is typically used to estimate the benefits of currency unions[2]. The last two variables are policy variables: RTA is a dummy variable for the presence of any trading agreement at time t and $tariff$ is the tariff that country j (importer) imposes on trade from country i at time t.

The gravity equation, either the simple version or its derived form, has been frequently used in various spheres of international trade. Researchers have used the gravity equation to explain the relationship between trade flows and other variables. Researchers use a gravity equation because it is a simple way to calculate welfare gains from trade. Head & Mayer (2014) emphasize that the

[2]See, e.g., Polák (2019) for the largest meta-analysis on euro currency union.

elasticity of trade or the import ratio are sufficient statistics (when their macro restrictions hold) for such calculation, and these values can be estimated by applying the gravity equation to international trade.

4.4 Dataset

Data collection is a crucial part of every meta-analysis because the quality of the data impacts the quality of the results. We use the results of primary studies for our meta-analysis, and for simplicity, we further use term 'trade cost' for the 'elasticity of trade with respect to bilateral trade cost'. As mentioned earlier, because of the popularity of the gravity equation in recent years, interest in analysing international trade has been increasing, which has enabled us to obtain an extensive dataset. To the best of our knowledge, a meta-analysis that properly investigates the studies related to the elasticity of trade with respect to trade costs has not been conducted, although Head & Mayer (2014) indirectly suggest such a study when discussing the policy impact of the elasticity of trade with respect to trade costs in their Cookbook. They build on the dataset from Disdier & Head (2008) and collect almost 160 studies estimating trade effects, but relatively few papers estimate trade cost elasticities.

The data for this study are collected from all studies on the trade costs effect that we could find. We started with the update of the dataset created by Disdier & Head (2008). We not only coded new variables but also verified the validity of the dataset itself. Since the dataset is more than 10 years old, we searched for new studies to enlarge the sample using RePEC database and Google Scholar to obtain both published and unpublished studies. We used combinations of the following keywords, 'gravity', 'trade cost', 'trade and quota', 'non-tariff', 'NTB', 'freight cost', 'freight charge', 'transit cost', 'shipment cost', 'tariff', and we searched titles, keywords and abstracts. We stopped the search of primary studies in September 2019 and did not include any new studies after that date. We excluded those studies with no empirical parts (from which we could not obtain any estimates for the meta-analysis) and downloaded those that might estimate trade costs.

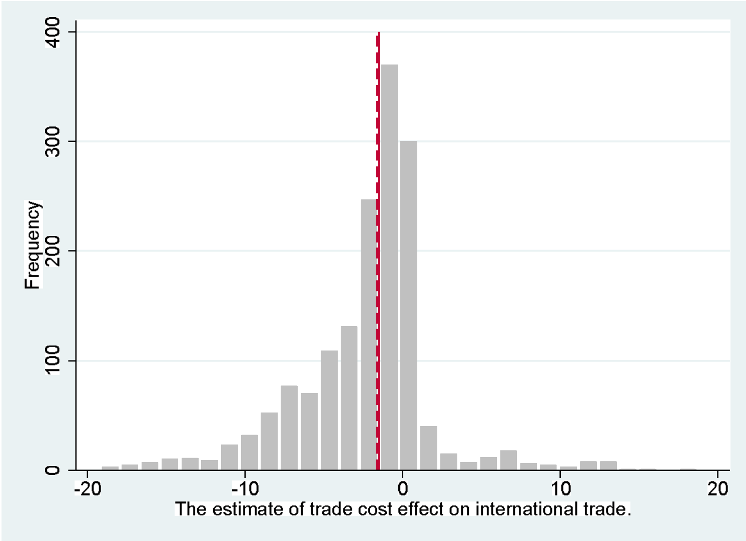
To create a homogeneous dataset suitable for meta-regression analysis, we applied the following criteria. (a) We included only empirical studies using the gravity framework presented in the Section 4.3 to analyse the trade costs. (b) We excluded all estimates analysing only bilateral trade within a country since our study is focused on international trade. (c) We included only studies written in English and German, as these were our language limitations. (d) We included only estimates with a reported estimation precision (standard error, p-value or t-statistic) because we need the precision in the form of the standard error to know how important the estimate should be in the meta-regression analysis and also for publication bias investigation. We collected all estimates from every study to obtain the largest possible dataset and to remain as unbiased as possible. Some authors highlight their preferred estimate, but in many cases, it is impossible to select a single estimate per study without making additional assumptions (e.g., precision, model or method suitability).

Outliers in the terms of trade elasticity and standard error are treated with method of winsorizing at 5% level., the final dataset contains 1609 estimates from 71 studies. Only 21 studies are working papers and unpublished studies; the majority of the papers are published in refereed journals, 22 in top 100 economic journals according to Repec. The list of the studies included in the dataset is available in Section 4.A. The majority of the gravity estimates are reported by studies that use regression derived from the basic Equation 4.1. Primary studies, from which we collect the estimates, use different methods to estimate trade elasticity with respect to trade costs or exporter 'competitiveness', such as wage, exchange rate, prices and productivity. We code these differences to understand if they are a significant determinant of the final effect. These differences are coded using 54 variables. The data were collected by the coauthors of this paper independently; then, we randomly checked some of the data collected by the other author to minimize mistakes that might be caused by manual entry and interpretation of the data.

The overall mean reported trade cost elasticity estimate is -2.23, and the overall median estimate is -1.43. This difference is might be caused by studies that report many estimates (some studies report only 1 estimate, and maximum number of estimates in a study is 136). To compensate for the effect of large studies, we weight the estimates by the inverse of the number of estimates reported and, hence, give the same weight to each study. This procedure results in a mean trade cost elasticity of -1.61, which is close to the median. The weighted number is more informative because it is not driven by, e.g., robustness checks and large studies. We further show that the number is exaggerated because of publication bias and that the true trade cost elasticity is much closer to zero.

Figure 4.2 depicts the histogram of the collected estimates of trade cost elasticities. The histogram shows that the majority of the estimates are negative:

Figure 4.2: Distribution of trade cost effects



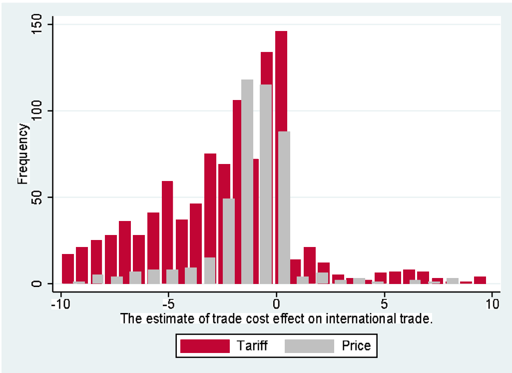
Note: The figure represents a histogram of the estimates of the trade cost coefficient reported in the individual studies without outliers (estimates with an absolute value greater than 20). The solid vertical line indicates the median of all the estimates. The long-dashed line indicates the median of the median estimates from individual studies.

only approximately 17% of the estimates are positive (which is theoretically implausible). One-third of the estimates are very close to zero – within the interval (-1,1). Some extreme outliers are included in the data; therefore, we excluded values with an absolute value greater than 20 from the chart. The elasticity varies not only between studies but also within studies. To understand the heterogeneity, we collected 54 variables to capture the differences between estimates.

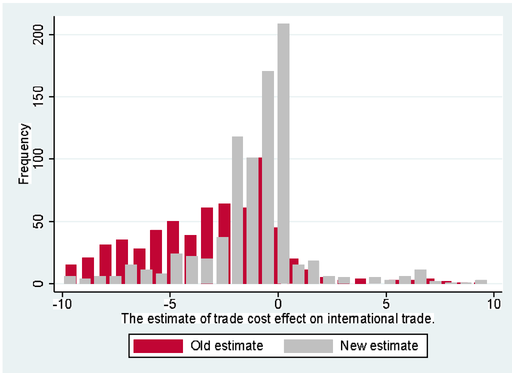
Systematic differences among the collected estimates were identified when we created different groups of estimates: some of them are shown in 4.3. These groups are created by characteristics that we code in the dataset and use to explain the drivers of variance among reported estimates. For instance, there is a difference between estimates for which different types of explanatory variables are used – the mean of the estimates based on tariffs is double the mean of estimates based on prices. The mean of the estimates based only on import data is -2.1, while the mean of the estimates based on export data is -1.0 (less than half). The mean of the estimates from panel series data is -2.0, while the mean for cross-sectional data is -2.6. Somewhat interesting is the difference in estimates prior to 2014 (those collected by Head & Mayer (2014)) and after that point. The mean of these "new estimates" is only -1.0, while the mean for

Figure 4.3: Key patterns in the data

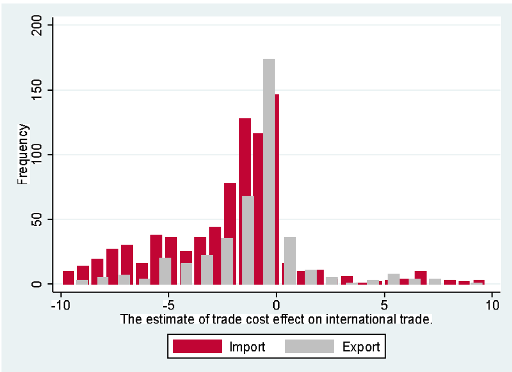
(a) Key variable



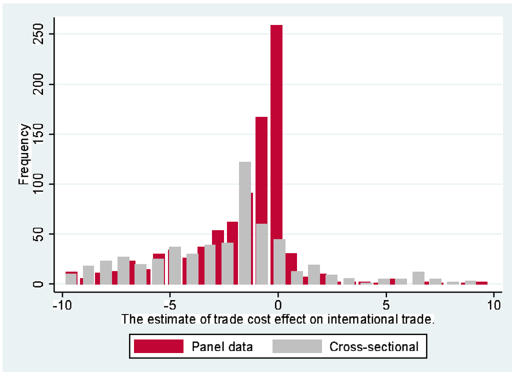
(b) New vs. old estimates



(c) Dependent variable



(d) Data dimension



Notes: Estimates smaller than -10 and larger than 10 are excluded from the figures for more representative histograms, but the estimates are included in all statistical tests and empirical estimation. Old estimates are those collected by Head & Mayer (2014); new estimates are from other studies. The variables shown here are chosen by expert judgment to illustrate the differences in data; not all of them have been proven to be significantly different from each other.

the "old estimates" is −3.8. These patterns in the data might help to explain the heterogeneity, but we cannot be sure if these differences are correlated with other factors or are fundamental. We investigate the heterogeneity using quantitative methods in Section 4.6 with collected variables.

Table 4.1 presents all the variables that our dataset captures from primary studies and provides variable definitions and basic summary statistics. We follow previous studies such as Polák (2019) and Havranek *et al.* (2015) in the variable definitions and variable grouping for better understanding. The collected variables should allow us to determine the drivers of heterogeneity and how choices made by researchers influence the final results and to address the most known methodological issues with the gravity equation, as described by Baldwin & Taglioni (2007) or Head & Mayer (2014).

Table 4.1: Description of regression variables

Variable	Description	Mean	SD
trade cost	"The estimate of trade cost effect on the international trade"	-2.23	4.21
se	Estimated standard error of trade cost.	0.86	1.04
price	=1 if the trade cost effect using prices not tariffs.	0.29	0.45
Data characteristics			
Mid year	The mid year of the sample on which the gravity equation is estimated (base is the sample minimum: 1870).	1994	16.62
Panel data	=1 if panel data are used in the gravity equation.	0.6	0.49
Dissagregated	=1 if data are disaggregated.	0.57	0.49
No. of obs.	The logarithm of the number of observations included in the gravity equation.	9.2	2.8
No. of years	The logarithm of the number of years.	1.4	1.3
Countries surveyed			
US	=1 if the trade cost effect is estimated for the US (or combinations of country groups)	0.56	0.5
Canada	=1 if the trade cost effect is estimated for Canada (or combinations of country groups)	0.51	0.5
Japan	=1 if the trade cost effect is estimated for Japan (or combinations of country groups)	0.61	0.49
EU	=1 if the trade cost effect is estimated for the EU (or combinations of country groups)	0.51	0.5
OECD	=1 if the trade cost effect is estimated for OECD countries.	0.31	0.46
Emerging	=1 if the trade cost effect is estimated for transition or developing countries.	0.41	0.49
Sector examined			
Agriculture	=1 if the trade cost effect is estimated for agriculture sector.	0.5	0.5
Animals	=1 if the trade cost effect is estimated for animal husbandry.	0.28	0.49

Continued on next page

Table 4.1: Description of regression variables (continued)

Variable	Description	Mean	SD
Transport	=1 if the trade cost effect is estimated for transport industry.	0.39	0.49
Service	=1 if the trade cost effect is estimated for service sector.	0.3	0.46
Raw Materials	=1 if the trade cost effect is estimated for raw materials.	0.37	0.48
Manufacture	=1 if the trade coes teffect is estimated for manufacturing or production sector.	0.67	0.47
Specification characteristics			
Model	=1 if the model is based on naive and not structural gravity	0.09	0.29
Total trade	=1 if the researchers sum export and import trade flows before taking logarithm form	0.2	0.4
Imports	=1 of the researchers measure only import flows	0.56	0.5
Exports	=1 of the researchers measure only export flows	0.29	0.45
No internal trade	=1 if within-country trade flows is estimated using production data	0.24	0.42
Asymmetry	=1 if the estiamte measures the international trade flows in one direction	0.44	0.5
Instruments	=1 if the instrument variable is used to correct for the endogeneity of GDP	0.17	0.37
Data source			
TRAINS	=1 if the researchers used a data from TRAINS database.	0.14	0.35
COMTRADE	=1 if the researchers used a data from COMTRADE database.	0.22	0.41
WITS	=1 if the researchers used a data from WITS database.	0.19	0.38
IMF	=1 if the researchers used a data from IMF database.	0.07	0.25
Treatment of multilateral resistance			
Control for MR	=1 if the gravity equation count for multilateral resistance terms.	0.54	0.5
Remoteness	=1 if the remoteness term is included.	0.23	0.42
Openess	=1 if the openess of the economy is included.		
OLS	=1 if the gravity equation is not estimated by OLS (reference category: OLS).	0.29	0.45
CFE	=1 if destination fixed effect is included.	0.52	0.5
Year FE	=1 if year fixed effect is included.	0.3	0.46
Sector FE	=1 if sector fixed effect is included.	0.39	0.49
Ratios	=1 if trade flow is normalized by trade with self.	0.18	0.38
Treatment of zero trade flows			
Zero omitted	=1 if observations of zero trade flows are omitted.	0.18	0.38
Zero plus one	=1 if the researchers add one to overall observations of zero trade flows.	0.08	0.26
PPML	=1 if the researchers applied Pseudo Poisson Maximum Likelihood method to estimate the gravity equation.	0.1	0.3

Continued on next page

Table 4.1: Description of regression variables (continued)

Variable	Description	Mean	SD
Control variables			
FTA	=1 if the gravity equation controls for free trade agreements.	0.25	0.43
Language	=1 if the gravity equation controls for common language.	0.44	0.5
Adjacency	=1 if the gravity equation controls for adjacency.	0.33	0.5
Distance	=1 if the gravity equation measure also the distance effect on the international trade.	0.53	0.5
Actual	=1 if the actual road/sea distance is used instead of great-circle formula.	0.16	0.37
Publication characteristics			
journal	=if the study is published in top ranked journals	0.3	0.46
isWP	=1 if the study is a working paper (reference category: published in a peer-reviewed papers)	0.33	0.47
firstpub	Year when the study first appeared in Google Scholar. (base 1985)	25.37	7.57
impact	Recursive discounted RePEc impact factor of the journal (collected in October 2019).	3.14	3.27
lnyearcits	Log of the mean number of Google Scholar citations (collected in October 2019).	1.95	1.94
new estimate	=1 if the estimate is not part of Head and Mayer (2014) dataset	0.56	0.47

Notes: SD = standard deviation. All variables except for citations and the impact factor are collected from studies estimating the trade cost (the search for studies was terminated on September 1, 2017, and the list of studies is available in Appendix). Citations are collected from Google Scholar, and the impact factor is collected from RePEc. Values are rounded to the closest hundredth.

Data characteristics The dataset directly impacts the final results; therefore, controlling for the dataset characteristics helps to understand the heterogeneity of the final results. Larger datasets should provide more precise estimates of the final effects. Thus, we calculate the logarithm of the number of observations in the dataset and the number of observations per year. Dummy variable 'panel' indicates whether the source dataset has panel characteristics. One might think that only panel data should be used to estimate trade costs, but only approximately half of the estimates use panel data. Some of our estimates use long time frames, and due to reduction in trade costs in the past, we wonder if the trade costs vary in time. To answer this question, we code the midpoint of the sample. Following Anderson & Van Wincoop (2004), who indicate that aggregated data yield different results than granular data when used in the gravity equation, we use a dummy that is equal to 1 if data are disaggregated at the sector or country level.

Data source The data source is one obvious possible source of heterogeneity, and we use dummy variables to control for different data sources: The UNCTAD Trade Analysis Information System (TRAINS); The United Nations Commodity Trade Statistics Database (UN Comtrade); and International Monetary Fund databases (IMF). These databases are used for the majority of the estimates: the rest are calculated using data from the OECD and other databases.

Countries surveyed The next category for variables is coding analysis of trade costs using data from different regions and countries. One reason is the general experience of obtaining different effects when analysing developed and emerging countries or more integrated groups of countries. We include several dummy variables on the basis of the frequency of the exploration within primary studies: the US, the EU, the OECD, Japan, Canada and emerging countries (including both transitioning and developing economies). The reference category for the dummies is remaining countries and their combinations or the entire world.

Sectors examined In the data characteristics section, we mentioned that researchers also use granular data. Some primary studies estimate the trade costs for different sectors of the economy, as these costs may vary substantially. The authors of primary studies generally follow two different approaches. The first approach is either to not consider sectoral heterogeneity at all or to consider only one of many sectors. The second approach is to disaggregate the effect of trade costs by sector (e.g., Caliendo & Parro (2014)). We decided to define dummy variables for the following sectors: agriculture, animals, transport, services, manufacturing and raw materials.

Specification characteristics We collected estimates from studies that estimate trade costs using the gravity equation. Results from other frameworks would not be easily comparable. Baldwin & Taglioni (2007) describe the most obvious mistakes that researches commit when applying the gravity framework or making methodological adjustments. The most well known is the 'silver medal mistake', where a simple sum or average of import and export is calculated before taking the logarithm [3]. Furthermore, following Polák (2019), we test whether models based on exports perform better than those based on imports, as suggested by Head & Mayer (2014).

Treatment of multilateral resistance Multilateral resistance has been discussed since Head & Mayer (2000) included a remoteness term to control for the distance between trade partners, but the issue was highlighted by Anderson & Van Wincoop (2003), who proposed a non-linear estimation method to control for multilateral resistance. Baldwin & Taglioni (2007) note that multilateral interactions between countries have positive effects on trade by raising the interest of firms to enter new markets; authors should control for such interactions when investigating international trade development. Several methods that use dummy variables for these effects are discussed in the literature. Another remedy to multilateral resistance (used even un-intentionally) is the time-varying fixed effects method for the gravity equation introduced by Harrigan (1993). More than half of the primary studies apply some type of fixed effects: country fixed effects capture systematic differences in the financial environment across countries (for example, bankruptcy laws) and sector fixed effects (issuing various kinds of industries) control for systematic differences in risk and performance across sector types. Variable year fixed effects control for differences in bilateral trade across years (as many authors consider panel data).

Control variables Control variables should be part of each gravity equation and typically include dummies for common language, common border and free trade agreement, but many other variables have been considered. Some of these variables cannot be included in the regression if they are constant for country pairs and a specific fixed effects method of estimation is used. We use

[3] Since the gravity equation is always estimated in logs, the authors use the log of the sum of bilateral trade as the response variable instead of the sum of the logs because the log of the sum (incorrect method) overestimates the sum of the log (correct method).

'0' for FTA and Language if these control variables are omitted even though they could be included. For distance, we control for not only its inclusion in the model but also the way it is measured - whether actual distance or the great-circle formula, which measures the distance between the centres of two countries, is used. The latter is easier to calculate but, according to Head & Mayer (2014), leads to upward bias in the estimated effect.

Treatment of zero trade flows If two examined regions do not trade with each other, a zero value is present in the dataset. This might appear to be acceptable, but since trade volume is a dependent variable, such observations can cause problems for the log-linear form of the gravity equation. The simplest solution is either to drop these zero trade observations or to replace zeros with ones. Silva & Tenreyro (2006) indicate that not considering zero observations causes significant bias when using OLS estimation. Other methods (e.g., Tobit model or Poisson pseudo maximum likelihood (PPML)) solve problems with zeros, and approximately 18% of the studies use these methods to avoid problems with zero trade values.

Publication characteristics Publication bias is an inseparable part of meta-analysis. We code the publication status of each study (working paper or published in refereed journal), and we control for journal quality following Polák (2017) using the RePEc database. We use RePEc Recursive Discounted Impact Factors, which reflect not only where the study was published but also the quality of citations that are split by their age in years (i.e., one for the current year). The main advantage of the RePEc database is that it covers nearly all economic journals and working paper series. Another publication characteristic is the number of citations on Google Scholar. We can test for a relation between the reported size of the trade cost effect and how such result is used. Year of publication can also be used to detect some trends in the results because new and more precise methods of estimation are developed or simply because people are becoming more efficient in every area, including international trade.

The variables we collected and described are intended to explain the variance in the gathered trade cost elasticities. We collected a total of 54 study, model and data characteristics, and a large number of possible combinations of these variables can be used in the regressions. We do not know a priori which ones we should include. We use Bayesian model averaging (BMA) to apply the most unbiased method for variable selection and follow recent trends in meta-analysis.[4]

4.5 Publication bias

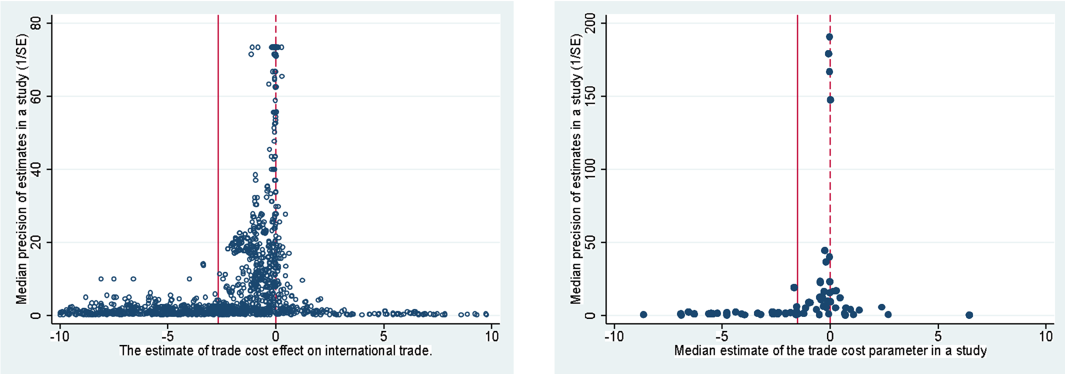
Testing for publication bias should be an essential part of every meta-analysis. Publication bias generally describes the situation in academic research where reported estimates do not correspond to the statistical prediction; that is, that the probability that an estimate is reported in a primary study is dependent on either the precision or magnitude of the estimate. Usually statistically insignificant results are under-reported, as are coefficients with unexpected or theory-inconsistent signs. This occurs if researches are motivated to prefer estimates that fulfil some criteria. For trade costs, the expected sign is negative – the higher the cost is, the lower the trade volume. Assume that true trade cost elasticity is negative: when we consider a large amount of different data sources and apply different methods, both positive and statistically insignificant estimates should still be present.

The most common and frequently used method for testing for publication bias in meta-analysis is the funnel asymmetry test (FAT); see, e.g., Stanley & Doucouliagos (2010) for more details. This method assesses the relationship between effect size and precision – usually the inverse of the standard error. A graphical illustration of this relationship is usually presented as a funnel plot, which is a scatter plot in which the size of the trade cost elasticity is on the horizontal axis and its precision is on the vertical axis. The test is based on the

[4]See, e.g., Polák (2019), Havranek *et al.* (2018b), Havranek *et al.* (2018a), Havranek & Irsova (2017)

Figure 4.4: Funnel plots

(a) All estimates (b) Median estimates for each study



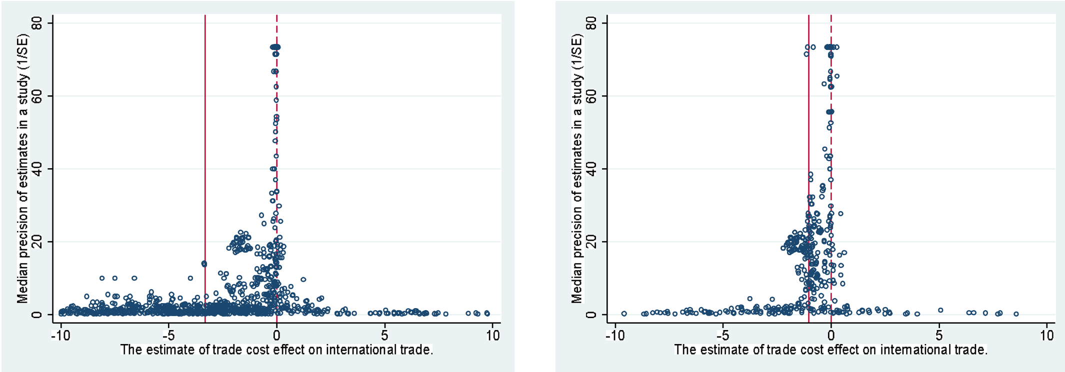
Notes: A funnel plot is a scatter plot where estimate (trade cost effect size) is plotted against precision (inverse of the standard error). The vertical solid line denotes the zero value, and the dashed line denotes mean value. Estimates smaller than -10 and larger than 10 are excluded from the figures for more representative plots, but the estimates are included in all statistical tests and empirical estimations.

shape of an inverted funnel that we would expect if no publication bias were present. The most precise estimates should be close to the true mean elasticity, and estimates further from the true value are expected to be imprecise, more dispersed and symmetrical around the mean. 4.4 shows two funnel plots. The left panel shows all collected estimates, and the right panel shows just one observation per study. In both cases, the most precise estimates are close to zero but far from the median value. The shape is not symmetrical; therefore, we should conduct a more precise test for the magnitude of the publication bias.

The funnel plots of our data are presented in 4.4. Panel (a) shows the funnel plot for the trade cost effect of all estimates, while panel (b) shows the funnel plot for the median estimates reported in the primary studies. The funnel plot is a graphical test and provides only an indication of possible bias that we should test for empirically. However, we can make several observations. First, the two funnels are not symmetrical and the estimates are not evenly distributed – the left side is much heavier. Second, the most precise estimates are distributed around zero (approximately one-third of all the estimates are close to zero), but these values are not close to the median or average. Third,

Figure 4.5: Funnel plots by identifying variables

(a) Tariff/freight rates (b) Exchange rates/producer prices



Notes: A funnel plot is a scatter plot where the estimate (trade cost effect size) is plotted against its precision (inverse of the standard error). The vertical dashed line denotes the zero value, and the solid line denotes the mean value. Estimates smaller than -10 and larger than 10 are excluded from the figures for more representative plots, but the estimates are included in all statistical tests and empirical estimations.

the funnel plot is not hollow. Last, the funnel with all the estimates appears to have more peaks (groups of precise estimates), which implies heterogeneity in the trade cost effect. We expect heterogeneity based on charts and statistics presented in 4.3.

The funnel comparison in 4.5 shows the difference between estimates based on different identifying variables (within the structural gravity model). Panel (a) depicts estimates that use tariffs/freight rates, and panel (b) shows those based on exchange rates/producer prices. These variables are used to estimate the same value but might lead to such heterogeneity that we should consider them separately (Head & Mayer, 2019). Both funnel plots are heavier on their left side than on their right side, which indicates the possible presence of publication bias caused by researchers who prefer more negative estimates and disregard positive ones.

The funnel plot helps to illustrate the gathered data and to form hypotheses about the presence of publication bias, but empirical tests are necessary to determine the magnitude of the bias. The main assumption is that methods used to estimate trade costs yield a symmetrical distribution; therefore, estimates and standard errors should be independent from a statistical perspective

(Stanley & Doucouliagos, 2010). If no bias exists, there would be no correlation between the estimates and their standard errors. Significant correlation, on the other hand, indicates the presence of publication bias. To test for funnel asymmetry and independence, we follow Stanley (2005) and use the following equation:

$$Tradecost_{ij} = Tradecost_0 + \beta SE\left(Tradecost_{ij}\right) + \epsilon_{ij}, \tag{4.7}$$

where $Tradecost_{ij}$ is the i-th estimate of the elasticity recorded on the j-th study, $SE\left(Tradecost_{ij}\right)$ denotes the standard errors of the elasticity estimate, $Tradecost_0$ is the mean elasticity corrected for possible publication bias, β measures the magnitude of the publication bias, and ϵ_{ij} is a normal disturbance term. Stanley (2005) noted that if the data are not affected by publication bias, the estimation effects will vary around the true effect $Tradecost_0$. If true elasticity was zero (signifying no trade cost effect) but 5% of the estimates are positive and statistically significant, the estimated β would be close to two, indicating that the researchers would require their t-statistics ($Tradecost/SE\left(Tradecost\right)$) to be at least two (Havranek *et al.*, 2015). Following the recommendation of Reed (2015) we use multiple estimation methods. We are also aware of the limitations raised by Alinaghi & Reed (2018) about this procedure but we also do not divide both sides of the regression with the standard error as e.g. Stanley *et al.* (2008).

The results of the test presented in Table 4.2 come from several specifications of Equation 4.7, and all are linear. Since the estimates from each study j are not likely to be completely independent, we use a clustering procedure to control for intra-study correlation. In all specifications (except the plain OLS), we find a statistically significant and negative coefficient for the standard error, which is interpreted as publication bias, and a significant and negative intercept, which is the mean trade cost elasticity corrected for publication bias. The overall mean elasticity in the data is -2.2, but after the correction, the value changes to -1.5 - -0.7. In any case, the effect is smaller, and this result is robust across all specifications.

Table 4.2: Funnel asymmetry test results for the whole dataset

	OLS	FE	BE	Precision	Study	IV
SE	-0.929**	-0.972***	-1.383***	-2.419***	-1.344***	-1.853***
(publication bias)	(-2.30)	(-5.62)	(-3.79)	(-4.64)	(-3.55)	(-2.64)
Constant	-1.616***	-1.575***	-0.769**	-0.260**	-0.797***	-0.727
(corrected mean)	(-4.88)	(-9.46)	(-2.10)	(-2.11)	(-3.50)	(-0.73)
Observations	1609	1609	1609	1609	1609	1609
rmse	3.288	2.438	2.203	0.961	2.821	3.470

Notes: OLS = ordinary least squares. FE = study-level fixed effect. BE = study-level between effects. Precision = the inverse of the reported estimate's standard error is used as the weight. Study = the inverse of the number of estimates reported per study is used as the weight. IV = the inverse of the square root of the number of observations used by researchers is used as an instrument for the standard error (the correlation between this instrument and the standard error is 0.3). Standard errors in parentheses. Standard errors are clustered at the study level. * p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01

The first column of Table 4.2 reports simple OLS regressions. The second column shows the fixed effect regression with study-level fixed effects to account for study-level characteristics. The third column uses between-study variance instead of within-study variance, which reduces the mean by half but almost doubles the magnitude of the publication bias. The next two columns use weights in the regression. Following Stanley & Doucouliagos (2017), the precision (inverse standard error) is used to account for the heteroskedasticity in the regression. The second type of weighting is by the inverse of the number of observations reported in a study, which gives each study the same weight in the final results – studies with more precise estimates or a greater number of estimates do not dominate the results. The last column builds on the instrumental variable (IV) approach, where the inverse of the square root of the number of observations is used as the instrument for the standard error. The IV approach is widely used in meta-analysis as a robustness check since the standard error is estimated jointly with the effect size in the primary studies. An intuitive instrument is based on the number of observations since a greater number of observations should lead to more precise estimates. Then, we consider several modifications (such as the square root, logarithm or count) to find the instrument with the largest correlation with the standard error. Our instrument is not ideal since the correlation is only 0.3. This approach does not provide a significant estimate of the mean trade cost elasticity and shows very strong publication bias.

Table 4.3: FAT-PET for different identifying variables

	OLS	FE	BE	Precision	Study	IV
A: Tariff						
SE	-0.839	-1.015***	-1.567***	-2.247***	-1.424***	-2.264***
(publication bias)	(-1.64)	(-4.75)	(-3.79)	(-3.86)	(-3.15)	(-3.13)
Constant	-2.040***	-1.844***	-0.899**	-0.477**	-1.018***	-0.459
(corrected mean)	(-4.19)	(-7.78)	(-2.03)	(-2.56)	(-3.27)	(-0.34)
Observations	1143	1143	1143	1143	1143	1143
rmse	3.575	2.582	2.222	1.584	2.944	3.977
B: Price						
SE	-0.866***	-0.849***	-0.180	-1.496***	-0.634	-0.142
(publication bias)	(-8.86)	(-4.30)	(-0.29)	(-4.26)	(-1.09)	(-0.19)
Constant	-0.861***	-0.871***	-0.721	-0.482**	-0.468**	-1.297**
(corrected mean)	(-3.47)	(-7.34)	(-1.57)	(-2.16)	(-2.21)	(-1.98)
Observations	466	466	466	466	466	466
rmse	2.250	2.040	1.743	0.730	2.302	2.377

Notes: OLS = ordinary least squares. FE = study-level fixed effect. BE = study-level between effects. Precision = the inverse of the reported estimate's standard error is used as the weight. Study = the inverse of the number of estimates reported per study is used as the weight. IV = the inverse of the square root of the number of observations used by researchers is used as an instrument for the standard error (correlation between this instrument and the standard error is 0.3). Standard errors in parentheses. Standard errors are clustered at the study level. * p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01

We previously noted that combining estimates of the trade cost that use price as the identifying variable and those that use tariffs might not be optimal. To explore the presence of publication bias in these two approaches, we estimate Equation 4.7 for these two subgroups of estimates separately and present the results in Table 4.3. In the key specifications, we again find a significant presence of publication bias and a corrected mean. The corrected mean for tariffs is approximately twice the size of that for prices – for tariffs, the mean trade cost elasticity is approximately -1.8 and for prices, it is -0.9 if we use the fixed effects model as the most reliable model and the rest of the specifications for robustness purposes. These results are in line with 4.3a and with the means of these two groups. The IV approach does not produce any significant results, but the instrument is weak.

4.6 Why Do Trade Cost Vary? – Modeling heterogeneity

In Section 4.4, we presented 4.3 and discussed several characteristics of primary study designs that might influence the reported estimates of trade cost elasticity. These characteristics are clearly not the only aspects: many others may play a crucial role. To model the heterogeneity, we collected 54 variables coding data characteristics, countries surveyed, examined sectors, model specification details, data source, approach towards multilateral resistance, treatment of zero trade flows and publication characteristics – see Table 4.1 for the full list. To model the relationship, we use the trade cost estimate as the dependent variable and all the characteristics as independent variables, which can be formally expressed in Equation 4.8 as extension of Equation 4.7:

$$Tradecost_{ij} = \alpha + \beta SE\left(TCelasticity_{ij}\right) + \gamma X_{ij} + \omega_{ij}, \tag{4.8}$$

where $TCelasticity_{ij}$ represents the i-th estimate of trade cost elasticity reported in study j, X_{ij} is a vector of the independent variables described in Section 4.4, and γ is the corresponding vector of estimates explaining the diversity of trade costs from primary studies. β represents the publication bias, and α (the intercept) is the mean elasticity corrected for publication bias that cannot be interpreted alone, only with the rest of the variables.

The large number of potential explanatory variables raises another issue – model uncertainty. Although we might have reasons for some of the variables to be important determinants of trade cost elasticity, we do not know for sure ex ante which variables are important and which study design characteristics do not influence the output and only add noise to the data. If we perform regression with all the variables, we would reduce the precision of our estimate by including considerable noise in the data. A common approach to address this problem is stepwise general-to-specific, a method for reducing the number of explanatory variables based on their level of significance. However, this process could exclude some important variables by accident. In recent years, increased computational power has led to the development of new model averaging techniques.

We use model averaging to address model uncertainty and minimize bias. We run a very large number of regressions with different combinations of variables and then aggregate the results of these models using weights. A great introduction to this approach is provided by Moral-Benito (2015), and we can further refer the reader to Steel (2019), who provides a survey of the most important methodological approaches to model averaging in economics and discusses the widely used BMA technique, which we also use to examine the heterogeneity among trade cost estimates.

BMA and frequency model techniques are suitable for running regressions on different sub-samples of a dataset to assess the robustness of each estimated coefficient. Each regression takes time, and the time complexity increases with the addition of each variable. To simplify the problem, we restrict the model space using BMA. BMA estimates a model and calculates the posterior model probability (PMP), which is an indicator of how well the model fits the data, similar to R^2. PMP is calculated for each estimated model, and the values are used as weights for the final estimated coefficients. BMA does not calculate the standard error for each coefficient but reports the posterior inclusion probability (PIP) for each coefficient, which represents the likelihood of inclusion in the final model. The closer PIP is to 1, the more relevant a variable is for the final model. The general details of BMA are described in several studies, e.g., Eicher *et al.* (2011) and Wright (2008). For the estimation itself, we use the BMA package in R software presented and described by Feldkircher & Zeugner (2009). BMA is also very frequently used among meta-analysts, see e.g. Meriluoto *et al.* (2019), Polák (2019), Gechert *et al.* (2019) for similar application.

Figure 4.6 graphically captures the BMA results: each row represents one explanatory variable. Variables are sorted by importance (by PIP value in descending order). Each column captures the results of one regression model, and the width of the column represents the PMP, sorted from highest value starting on the left. Cells capture the sign of the coefficients of each variable in a given model: a white cell indicates that the variable is not included, red colour indicates a negative sign and blue indicates a positive sign. The figure shows how stable (in terms of sign) each variable is across the estimated models.

In the previous part, we showed that publication bias is present in the trade cost literature, and the first conclusion we can draw from the BMA is that the variable capturing publication bias (the standard error of an estimate) is the variable with the highest PIP when modelling the heterogeneity in the reported estimates. Publication bias is a serious issue, and these results confirm that it is not driven by data characteristics or model properties that were not part of Equation 4.7, which we used in the previous part. However, publication bias is not the only key driver of heterogeneity. A few other variables have a PIP very close to 100%. The results of BMA are reported in Table 4.4, along with the results of an OLS regression based on Equation 4.8 for robustness purposes. For the robustness check, we used the variables selected by the BMA with a PIP greater than 50% and then performed OLS with clustered standard errors. The results are consistent, and all the used variables are significant at the 5% significance level.

Data source and characteristics Several variables that code the data characteristics used in the primary studies were found to be important drivers of the trade cost elasticity estimate. Our results suggest that panel data lower the elasticity – estimates using panel data show stronger effects of trade costs on trade. This finding is fully in-line with the key patterns found in the data and presented in 4.3d. Data disaggregation has the opposite effect on the estimate – more granular data (industry level or even firm level) leads to a smaller effect of trade costs (closer to zero). In other words, data for the whole economy (aggregated) indicate that trade costs have a large effect on trade. Furthermore, the greater the number of years included in the dataset is, the more positive the estimate.

Another key pattern in the data, the difference between estimates based on price or tariffs as the identifying variable captured by 4.3a, is not confirmed by the BMA analysis, and the PIP of this explanatory variable is very low.

Countries surveyed We do not have prior expectations about such variables, but of the six variables coding countries present in the dataset, three systematically influence the size of the trade cost elasticity estimates. If Canada is included in the dataset, the trade cost are more positive. If Japan is included, the trade costs are very negative. If all OECD countries are part of the dataset (approximately one-third of the collected estimates), the trade cost elasticity estimate is less positive. However, the PIP of this variable is only 0.7, which is at the edge of our threshold for variables that are considered to be important. For the dataset that includes all countries, the effects would not compensate for each other, so the selection of the countries used to estimate trade costs is important.

Specification characteristics One of the variables coding the differences in model specification related to data type was found to be a significant driver of heterogeneity in the trade cost elasticity estimates. We identified this variable previously when presenting the key pattern in the data – the difference between estimates using import and export data as the explanatory variable, captured by 4.3c. Data based on export flows imply more positive trade cost elasticity. Polák (2019) argues that export data are more reliable than import data or total trade data. We confirmed that this difference plays an important role in the estimation process and is a source of bias.

Another important specification characteristics is the use of country-level or country-pair fixed effects. This approach is one remedy for multilateral resistance (discussed below) and is also an important driver of heterogeneity. If country-level fixed effects are used, the trade cost elasticity estimate is more positive, indicating that trade costs do not substantially influence trade.

Control of zero trade flows Zero trade flows often occur in large datasets with many countries. Several methods can be used to address such data points, and the choice of method impacts the trade cost elasticity estimate. If zero trade flows are replaced with one, the results are more positive than if the zero trade flows are not considered. Notably, less than 10% of our trade cost elasticity estimates use this approach. The results are not significantly different when the PPML method is used to estimate the gravity equation.

Treatment of multilateral resistance Anderson & Van Wincoop (2003) describes why treatment of multilateral resistance is important when estimating the gravity equation. We identified several study properties that have a significant impact on the trade cost elasticity estimate. If the study controls for multilateral resistance, the estimates are more positive. The same effect is observed for normalization of trade flows with trade itself (gravity is estimated using ratios). Another approach is to use remoteness terms, but this method does not affect the final results of the primary studies.

Control variables The gravity equation provide a large number of control variables, but some cannot be used if a large number of fixed effects are used. Several variables were found to have a significant impact on the trade cost elasticity estimate. More positive results are obtained from models that include controls for FTA and country openness. Country openness is rarely included, but Balavac & Pugh (2016) demonstrate the importance of country openness and terms of trade with respect to trade volumes. The basic control variable is the distance between countries, but not every researcher includes distance. If distance is controlled for, the trade cost elasticity is more negative. The method used to measure distance does not matter.

Publication characteristics Although we identified serious publication bias, the publication characteristics we collected do not contribute to the heterogeneity of the estimates. No difference is observed between published studies or working papers or between results published in top economics journals and those published elsewhere. Interestingly, the estimate of trade cost elasticity does not depend on the year in which the study was published (suggesting that there is no change in trade cost elasticity over time). On the other hand, the only publication characteristic that was found to be important belongs to the patterns in the data – the difference between new and old estimates, as shown in 4.3b. The estimates included in Head & Mayer (2014) are more negative than estimates from studies published after that date. If we further consider the fact that datasets that include more years of data produce less negative trade costs elasticity estimates, our results indicate that the effect of trade costs is decreasing over time.

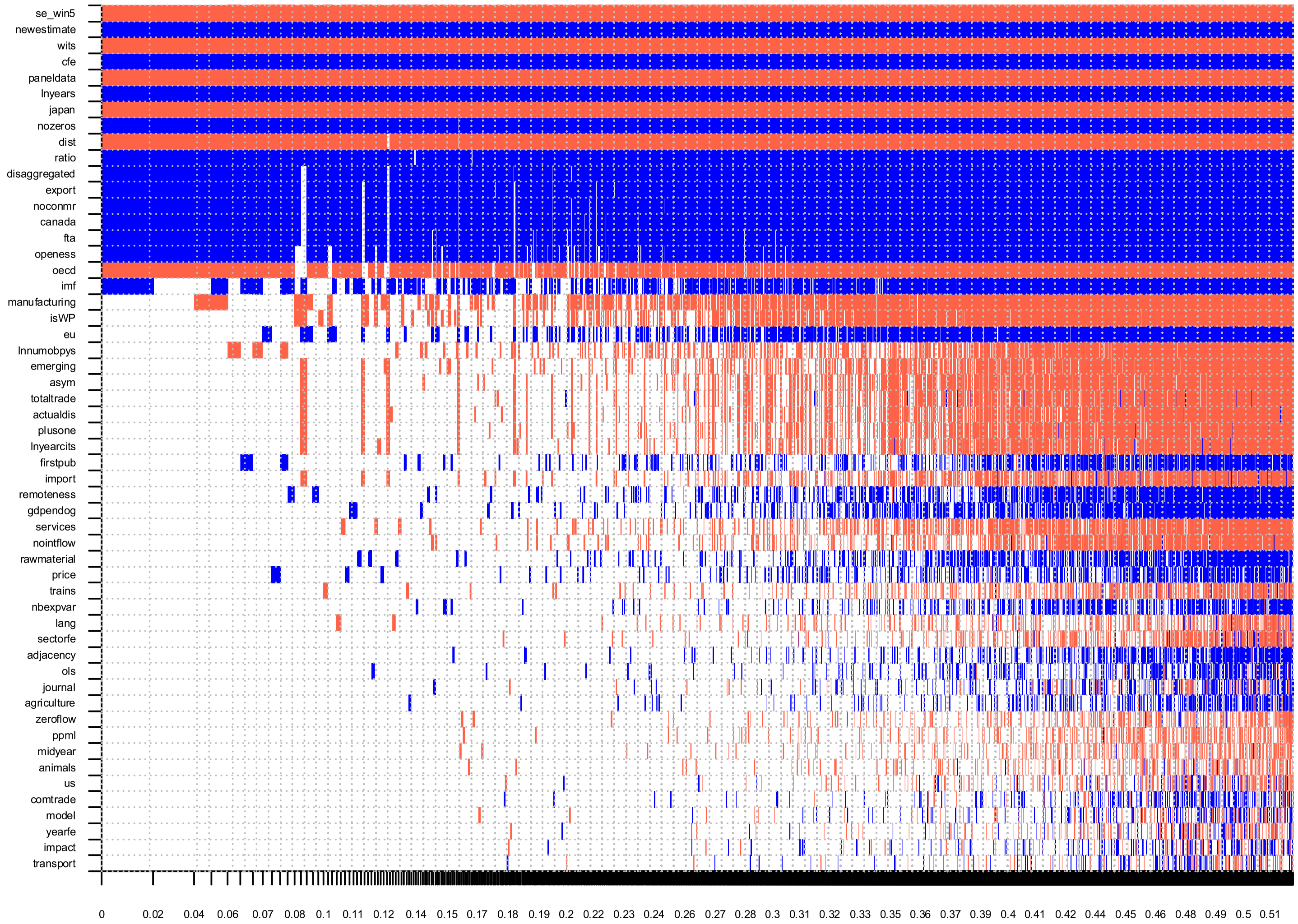
4.6.1 Best practice implied elasticity

In the next paragraphs, we attempt to estimate the mean implied elasticity of trade costs given the best practices. This approach creates a synthetic study in which all the collected data are used, but we give different weights according to the best practice methodology using a very large number of observations and the most recent data, following previous meta-analyses. The definition of "best practices" is subjective since the best approach is unclear for some of the variables. For each variable presented in Table 4.1, a preferred value is selected or left unchanged. We use these results to calculate the implied elasticity.

For each variable presented in Table 4.1, a preferred value is selected (or the value is left unchanged for a given estimate if the study does not have a preference for the value of the variable), and the implied elasticity is computed as the mean predicted estimate of elasticity. This approach creates a synthetic study with the best practice methodology using a very large number of observations and the most recent data, following previous meta-analyses.

The results are presented in Table 4.5. The overall mean elasticity is reported in the last row. The column labelled "Difference from mean" shows the difference between the calculated predictions and the simple means described in Section 4.4. The results vary substantially because of the heterogeneity present in the data – the overall mean elasticity of approximately -0.63 is not significantly different from zero. The precision of our best practice estimate reflects the uncertainty in the estimates of regression parameters of the OLS

Figure 4.6: Model inclusion in Bayesian model averaging



Notes: These are the estimation results of equation 4.8 based on best 5000 models. Response variable: t-statistic of the euro coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one). All moderators are divided by the standard error of the euro coefficient estimate. Columns denote individual models, and variables are sorted by posterior inclusion probability in descending order from the top. Blue colour (darker in grayscale) = the variable is included and the estimated sign is positive. Red colour (lighter in grayscale) = the variable is included and the estimated sign is negative. White colour = 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 4.4. A detailed description of all variables is available in Table 4.1.

Table 4.4: Explaining the differences in the estimates of the trade costs

Response variable:	Bayesian model averaging			OLS		
Estimate of τ	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value
se (publication bias)	-0.96	0.069	1	-0.966	0.154	0
price	0.019	0.096	0.054			
new estimate	2.794	0.454	1	2.912	0.337	0
Data characteristics						
Mid year	0	0.001	0.018			
Panel data	-2.578	0.378	1	-2.746	0.625	0
Dissagregated	0.73	0.253	0.951	0.781	0.396	0.053
No. of obs.	-0.013	0.034	0.161			
No. of years	0.738	0.144	0.999	0.788	0.241	0.002
No. Od vars.	0.001	0.007	0.038			
Countries surveyed						
US	-0.004	0.072	0.015			
Canada	1.222	0.545	0.887	1.58	0.478	0.002
Japan	-1.898	0.404	0.993	-2.088	0.399	0
EU	0.204	0.374	0.28			
OECD	-0.681	0.506	0.702	-0.996	0.453	0.031
Emerging	-0.079	0.229	0.131			
Sector examined						
Agriculture	0.004	0.038	0.02			
Animals	-0.003	0.039	0.015			
Transport	0	0.02	0.011			
Service	-0.037	0.14	0.087			
Raw Materials	0.024	0.104	0.066			
Manufacture	-0.197	0.27	0.402			
Specification characteristics						
Model	0	0.05	0.013			
Total trade	-0.215	0.664	0.114			
Imports	-0.091	0.314	0.094			
Exports	1.345	0.484	0.914	1.571	0.308	0
No internal trade	-0.041	0.179	0.068			
Asymmetry	-0.111	0.331	0.125			
Instruments	0.043	0.163	0.087			
Data source						
TRAINS	-0.018	0.106	0.041			
COMTRADE	0.001	0.034	0.014			
WITS	-2.068	0.249	1	-2.079	0.323	0
IMF	0.591	0.723	0.463			
Treatment of multilateral resistance						
Control for MR	0.957	0.387	0.908	1.069	0.3	0.001
Remoteness	0.049	0.18	0.089			
Openess	0.665	0.443	0.753	0.932	0.536	0.087
OLS	0.005	0.045	0.023			
CFE	1.204	0.24	1	1.14	0.273	0
Year FE	-0.01	0.089	0.026			
Sector FE	-0.001	0.034	0.013			
Ratios	0.959	0.337	0.973	0.948	0.41	0.024
Treatment of zero trade flows						
Zero flow in the data	-0.004	0.045	0.019			
Zero omitted	1.521	0.395	0.981	1.713	0.381	0
Zero plus one	-0.104	0.328	0.112			
PPML	-0.004	0.052	0.018			
Control variables						
FTA	0.85	0.416	0.862	1.137	0.265	0
Language	-0.009	0.068	0.03			
Adjacency	0.008	0.064	0.026			
Distance	-1.489	0.37	0.98	-1.656	0.29	0
Actual	-0.1	0.322	0.113			
Publication characteristics						
journal	0.001	0.065	0.021			
isWP	-0.237	0.422	0.296			
firstpub	0.005	0.018	0.095			
impact	0	0.004	0.013			
lnyearcits	-0.036	0.119	0.104			
Constant	-2.851	NA	1	-3.544	0.637	0
Studies	71			71		
Observations	1,609			1,609		

Notes: Response variable: t-statistic of the euro coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for). All moderators are divided by the standard error of the euro coefficient estimate except for publication year, which is interacted with SE in the original model. PIP = posterior inclusion probability. SD = standard deviation. VIF = variance inflation factor. A detailed description of all variables is available in 4.1.

model. However, even higher uncertainty might be related to the definition of the best practice values of several variables, which would make the confidence intervals even wider.

Table 4.5: Best practice predictions of trade cost elasticity

Sample	Estimate	95% conf. int.		Difference from mean
Old studies	-2.154	-4.328	3.887	1.753
New studies	0.640	-2.205	6.006	1.993
Full sample	-0.626	-3.098	4.986	1.884

Notes: This table presents the estimates of the trade cost coefficient for selected groups implied by BMA and OLS and our definition of best practice. That is, we take the regression coefficients estimated by BMA (Table 4.4) and predict the values of *trade cost* conditional on aspects of the methods and data. Difference from mean = the difference between these estimates and the simple means reported in Table 3.1. The confidence intervals are approximate and constructed using the standard errors estimated by OLS.

4.7 Conclusion

International trade is currently under pressure from increased protectionist measures taken by large world economies. Economists have been investigating the factors influencing international trade for decades. The most commonly used gravity framework has become a standardized tool for trade cost analysis. From a theoretical perspective, any barriers to trade would harm trade. The question is how serious the tariffs and price measures imposed on traded goods are. To answer this question, we use meta-analytic techniques to analyse over 1600 trade cost elasticity estimates from over 70 studies that used the gravity framework. These estimates range from -78.6 to 77.0, with a mean of -2.66 and median of -1.52. To explain this substantial heterogeneity, we collected an additional 54 characteristics related to each estimate that represent the precision, the study and model design, dataset source and properties, control variables used in the estimation process and the characteristics of the publication itself and its impact in the academic community.

We build upon Head & Mayer (2014), who focused quantitatively on the elasticity of trade with respect to trade costs and made the first attempt to aggregate the findings. The authors computed mean and median estimates of the trade cost coefficients and compared the overall effect with separate summary statistics for particular groups created based on study properties (type of gravity equation, type of identifying variable). We further extended the dataset by collecting additional studies and explanatory variables. In addition, we focused on whether possible publication bias could lead to exaggerated results.

Trade wars are bad for international trade, but the literature suffers from publication bias – estimates are tuned or selected to be statistically significant and in-line with economic theory. The results of graphical tests were supported by empirical methods (funnel asymmetry test), which estimated the size of the underlying effect beyond publication bias to be approximately -1.4, close to the median value. Publication bias was also significant in explaining the heterogeneity among the estimates. Publication bias is an issue, but we did not identify any publication characteristics, such as the quality of the journal or when the study was published, that systematically influenced trade cost elasticity.

We identified several properties that have a significant impact on the estimated size of trade costs. From a methodological perspective, how zero trade flows are addressed and whether the study uses country fixed effects are important factor. The dataset is also important – there are differences between estimates based on import and export data and estimates from different data sources. Due to the data heterogeneity, conditional estimates (based on the preferred combination of explanatory variables) of trade elasticity have such large confidence intervals that the results are not significantly different from zero. Therefore, we cannot make any conclusions based on these data aggregates.

Bibliography

Adam, M. C. (2019): "Return of the tariffs: The interwar trade collapse revisited." *Freie Universität Berlin, School of Business & Economics Discussion Paper* **(8)**.

Ahn, J., A. K. Khandelwal, & S.-J. Wei (2011): "The role of intermediaries in facilitating trade." *Journal of International Economics* **84(1)**: pp. 73 – 85.

Alinaghi, N. & W. R. Reed (2018): "Meta-analysis and publication bias: How well does the fat-pet-peese procedure work?" *Research synthesis methods* **9(2)**: pp. 285–311.

Anderson, J. E. (1979): "A theoretical foundation for the gravity equation." *The American Economic Review* **69(1)**: pp. 106–116.

Anderson, J. E., I. Borchert, A. Mattoo, & Y. Yotov (2018): "Dark costs, missing data: Shedding some light on services trade." *European Economic Review* **105(C)**: pp. 193–214.

Anderson, J. E. & D. Marcouiller (2002): "Insecurity and the pattern of trade: An empirical investigation." *The Review of Economics and Statistics* **84(2)**: pp. 342–352.

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.

Arkolakis, C., A. Costinot, & A. Rodríguez-Clare (2012): "New trade models, same old gains?" *American Economic Review* **102(1)**: pp. 94–130.

Asche, F., H.-M. Straume, & E. Vårdal (2018): "Perish or prosper: Trade patterns for highly perishable products." *Working Papers in Economics 4/18*, University of Bergen, Department of Economics.

Baier, S. L. & J. H. Bergstrand (2001): "The growth of world trade: tariffs, transport costs, and income similarity." *Journal of International Economics* **53(1)**: pp. 1 – 27.

Bajzik, J., T. Havranek, Z. Irsova, & J. Schwarz (2019): "Estimating the Armington Elasticity: The Importance of Data Choice and Publication Bias." *Working Papers IES 2019/19*, Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Balavac, M. & G. Pugh (2016): "The link between trade openness, export diversification, institutions and output volatility in transition countries." *Economic Systems* **40(2)**: pp. 273–287.

Baldwin, R. & D. Taglioni (2007): "Trade effects of the euro: A comparison of estimators." *Journal of Economic Integration* pp. 780–818.

Bas, M., T. Mayer, & M. Thoenig (2017): "From micro to macro: Demand, supply, and heterogeneity in the trade elasticity." *Journal of International Economics* **108**: pp. 1–19.

Békés, G., L. Fontagné, B. Muraközy, & V. Vicard (2012): "How frequently firms export? evidence from france." *IEHAS Discussion Papers MT-DP - 2012/17*, Budapest.

von Below, D. & P.-L. Vézina (2016): "The trade consequences of pricey oil." *IMF Economic Review* **64(2)**: pp. 303–318.

Bergstrand, J. H. (1985): "The gravity equation in international trade: Some microeconomic foundations and empirical evidence." *The Review of Economics and Statistics* **67(3)**: pp. 474–481.

BERGSTRAND, J. H. (1989): "The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade." *The Review of Economics and Statistics* **71(1)**: pp. 143–153.

BERMAN, N. (2009): "Financial Crises and International Trade: The Long Way to Recovery." *Economics Working Papers ECO2009/23*, European University Institute.

BERMAN, N., P. MARTIN, & T. MAYER (2012): "How do different exporters react to exchange rate changes?" *The Quarterly Journal of Economics* **127(1)**: pp. 437–492.

BESEDEŠ, T. & M. T. COLE (2017): "Distorted trade barriers: A dissection of trade costs in a "distorted gravity" model." *Review of International Economics* **25(1)**: pp. 148–164.

BLONIGEN, B. A. & W. W. WILSON (2010): "Foreign subsidization and excess capacity." *Journal of International Economics* **80(2)**: pp. 200 – 211.

BURSTEIN, A. T., J. C. NEVES, & S. REBELO (2003): "Distribution costs and real exchange rate dynamics during exchange-rate-based stabilizations." *Journal of monetary Economics* **50(6)**: pp. 1189–1214.

CALIENDO, L. & F. PARRO (2014): "Estimates of the Trade and Welfare Effects of NAFTA." *The Review of Economic Studies* **82(1)**: pp. 1–44.

CHEONG, J., D. W. KWAK, & K. K. TANG (2018): "The trade effects of tariffs and non-tariff changes of preferential trade agreements." *Economic Modelling* **70**: pp. 370–382.

CHEVASSUS-LOZZA, E., D. MAJKOVIC, V. PERSILLET, & M. UNGURU (2005): "Technical barriers to trade in the european union: Importance for the new eu members. an assessment for agricultural and food products." *Technical Report 24621*, European Association of Agricultural Economists. 2005 International Congress, August 23-27, 2005, Copenhagen, Denmark.

Cipollina, M., F. Demaria, & F. Pietrovito (2016): "Determinants of trade: the role of innovation in presence of quality standards." *Journal of Industry, Competition and Trade* **16(4)**: pp. 455–475.

Cipollina, M., D. Laborde Debucquet, & L. Salvatici (2013): "Do preferential trade policies (actually) increase exports? an analysis of eu trade policies." *2013 Annual Meeting, August 4-6, 2013, Washington, D.C. 150177*, Agricultural and Applied Economics Association.

Ćorić, B. & G. Pugh (2010): "The effects of exchange rate variability on international trade: a meta-regression analysis." *Applied Economics* **42(20)**: pp. 2631–2644.

Costinot, A., D. Donaldson, & I. Komunjer (2011): "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *The Review of Economic Studies* **79(2)**: pp. 581–608.

Coughlin, C. C. & D. Novy (2012): "Is the international border effect larger than the domestic border effect? evidence from us trade." *CESifo Economic Studies* **59(2)**: pp. 249–276.

De, P. (2010): "Global economic and financial crisis: India's trade potential and prospects, and implications for asian regional integration." *Journal of Economic Integration* **25(1)**: pp. 32–68.

De Bromhead, A., A. Fernihough, M. Lampe, & K. H. O'Rourke (2019): "When britain turned inward: The impact of interwar british protection." *American Economic Review* **109(2)**: pp. 325–52.

De Melo, J. & J.-M. Solleder (2018): "Barriers to trade in environmental goods: How important they are and what should developing countries expect from their removal." .

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.

Dutt, P. & D. Traca (2010): "Corruption and bilateral trade flows: Extortion or evasion?" *The Review of Economics and Statistics* **92(4)**: pp. 843–860.

Eaton, J. & S. Kortum (2002): "Technology, geography, and trade." *Econometrica* **70(5)**: pp. 1741–1779.

Egger, P. & M. Pfaffermayr (2003): "The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects." *Empirical Economics* **28(3)**: pp. 571–580.

Egger, P. & J. Průša (2016): "The determinants of trade costs: a random coefficient approach." *Empirical economics* **50(1)**: pp. 51–58.

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.

Emlinger, C., E. C. Lozza, & F. Jacquet (2006): "Eu market access for mediterranean fruit and vegetables: A gravity model assessment." *98th Seminar, June 29-July 2, 2006, Chania, Crete, Greece 10098*, European Association of Agricultural Economists.

Erkel-Rousse, H. & D. Mirza (2002): "Import price elasticities: reconsidering the evidence." *Canadian Journal of Economics/Revue canadienne d'économique* **35(2)**: pp. 282–306.

Estevadeordal, A., B. Frantz, & A. M. Taylor (2003): "The rise and fall of world trade, 1870–1939." *The Quarterly Journal of Economics* **118(2)**: pp. 359–407.

FELDKIRCHER, M. & S. ZEUGNER (2009): *Benchmark priors revisited: on adaptive shrinkage and the supermodel effect in Bayesian model averaging*. 9-202. International Monetary Fund.

FINK, C., A. MATTOO, & I. C. NEAGU (2005): "Assessing the impact of communication costs on international trade." *Journal of International Economics* **67(2)**: pp. 428 – 445.

FLACH, L. & F. UNGER (2018): "Quality and gravity in international trade." **(12602)**.

FONTAGNÉ, L., T. MAYER, & S. ZIGNAGO (2005): "Trade in the triad: how easy is the access to large markets?" *Canadian Journal of Economics/Revue canadienne d'économique* **38(4)**: pp. 1401–1430.

FRANCOIS, J. & J. WOERZ (2009): "Non-linear panel estimation of import quotas: The evolution of quota premiums under the atc." *Journal of International Economics* **78(2)**: pp. 181 – 191.

FUGAZZA, M. & A. NICITA (2011): "Measuring preferential market access." *MPRA Paper 38565*, University Library of Munich, Germany.

FULLER, K. & P. L. KENNEDY (2019): "A determination of factors influencing sugar trade." *International Journal of Food and Agricultural Economics (IJFAEC)* **7(1128-2019-561)**: pp. 19–29.

GECHERT, S., T. HAVRÁNEK, Z. HAVRÁNKOVÁ, & D. KOLCUNOVA (2019): "Death to the cobb-douglas production function." *Technical report*, IMF Working Paper.

GERVAIS, J.-P., B. LARUE, T. OTSUKI, M.-L. RAU, K. SHUTES, C. WIECK, & N. WINCHESTER (2011): "New data and analysis on non-tariff measures in agri-food trade." *2011 Annual Meeting, July 24-26, 2011, Pittsburgh, Pennsylvania 103730*, Agricultural and Applied Economics Association.

Harrigan, J. (1993): "Oecd imports and trade barriers in 1983." *Journal of international Economics* **35(1-2)**: pp. 91–111.

Havránek, T. (2010): "Rose effect and the euro: is the magic gone?" *Review of World Economics* **146(2)**: pp. 241–261.

Havranek, T., D. Herman, & Z. Irsova (2018a): "Does daylight saving save electricity? a meta-analysis." *Energy Journal* **39(2)**.

Havranek, T., R. Horvath, Z. Irsova, & M. Rusnak (2015): "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

Havranek, T. & Z. Irsova (2017): "Do borders really slash trade? a meta-analysis." *IMF Economic Review* **65(2)**: pp. 365–396.

Havranek, T., Z. Irsova, & T. Vlach (2018b): "Measuring the income elasticity of water demand: the importance of publication and endogeneity biases." *Land Economics* **94(2)**: pp. 259–283.

Hayakawa, K. (2013): "How serious is the omission of bilateral tariff rates in gravity?" *Journal of the Japanese and International Economies* **27**: pp. 81 – 94.

Hayakawa, K. (2014): "Bilateral tariff rates in international trade: finished goods versus intermediate goods." *International Economics and Economic Policy* **11(3)**: pp. 353–370.

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 (2014): "Gravity equations: Workhorse, toolkit, and cookbook." In "Handbook of international economics," volume 4, pp. 131–195. Elsevier.

Head, K. & T. Mayer (2019): "Brands in motion: How frictions shape multinational production." *American Economic Review* **109(9)**: pp. 3073–3124.

Head, K. & J. Ries (2001): "Increasing returns versus national product differentiation as an explanation for the pattern of u.s.-canada trade." *American Economic Review* **91(4)**: pp. 858–876.

Heid, B., M. Larch, & Y. Yotov (2017): "Estimating the effects of nondiscriminatory trade policies within structural gravity models." .

Hertel, T., D. Hummels, M. Ivanic, & R. Keeney (2007): "How confident can we be of cge-based assessments of free trade agreements?" *Economic Modelling* **24(4)**: pp. 611 – 635.

Hoekman, B. & A. Nicita (2008): "Trade policy, trade costs, and developing country trade." *Technical report*, The World Bank.

Hou, Y., Y. Wang, & H. Yilmazkuday (2017): "Gravity channels in trade." *Globalization and Monetary Policy Institute Working Paper* **(297)**.

Hugot, J., C. U. Dajud *et al.* (2016): "Trade costs and the suez and panama canals." *Technical report*.

Hummels, D. (1999): "Toward a Geography of Trade Costs." *GTAP Working Papers 1162*, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

Jacks, D. S., C. M. Meissner, & D. Novy (2010): "Trade costs in the first wave of globalization." *Explorations in Economic History* **47(2)**: pp. 127 – 141.

Jakab, Z. M., M. A. Kovács, & A. Oszlay (2001): "How far has trade integration advanced?: An analysis of the actual and potential trade of three central and eastern european countries." *Journal of Comparative Economics* **29(2)**: pp. 276 – 292.

Jaroensathapornkul, J. (2017): "Effects of non-tariff barriers on thai exports of key vegetables to asean countries." *International Journal of Trade and Global Markets* **10(1)**: pp. 3–12.

Keita, M. (2016): "Does ict development flatten the globe? evidence from international trade costs data." *American Journal of Trade and Policy* **2(3)**: pp. 153–160.

Kinzius, L., A. Sandkamp, & E. Yalcin (2018): "Trade protection and the role of non-tariff barriers." *Review of World Economics* pp. 1–41.

Lai, H. & D. Trefler (2002): "The gains from trade with monopolistic competition: Specification, estimation, and mis-specification." *Working Paper 9169*, National Bureau of Economic Research.

Lanati, M. (2013): "Estimating the elasticity of trade: the trade share approach." *LEM Papers Series 2013/10*, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.

Liapis, P. S. (2011): "Changing patterns of trade in processed agricultural products." *OECD Food, Agriculture and Fisheries Papers 47*, OECD Publishing.

Limao, N. & A. J. Venables (2001): "Infrastructure, geographical disadvantage, transport costs, and trade." *The World Bank Economic Review* **15(3)**: pp. 451–479.

Martinez-Zarzoso, I. & F. Nowak-Lehmann (2003): "Augmented gravity model: An empirical application to mercosur-european union trade flows." *Journal of Applied Economics* **6(2)**: pp. 291–316.

Matyas, L., L. Konya, & M. Harris (2004): "Modelling export activity of eleven apec countries, 1978 - 1997." *Applied Econometrics and International Development* **4(4)**.

Mazhikeyev, A. & T. Edwards (2013): "Central asian trade relations in the post-soviet era." *Discussion Paper Series 2013-02*, Department of Economics, Loughborough University.

Melo, O., A. Engler, L. Nahuelhual, G. Cofre, & J. Barrena (2012): "Does sanitary and phytosanitary regulation stringency affect developing countries exports? evidence from chilean fruit exports." *2012 Conference, August 18-24, 2012, Foz do Iguacu, Brazil 125298*, International Association of Agricultural Economists.

Meriluoto, L., K. K. Das, J. Duan, & W. R. Reed (2019): "Spillovers and exports: A meta-analysis." .

Moral-Benito, E. (2015): "Model averaging in economics: An overview." *Journal of Economic Surveys* **29(1)**: pp. 46–75.

Nahuis, R. (2004): "One size fits all?: Accession to the internal market; an industry-level assessment of eu enlargement." *Journal of Policy Modeling* **26(5)**: pp. 571 – 586.

Novy, D. (2006): "Is the iceberg melting less quickly? international trade costs after world war ii." *The Warwick Economics Research Paper Series (TWERPS) 764*, University of Warwick, Department of Economics.

Novy, D. (2013): "Gravity redux: measuring international trade costs with panel data." *Economic inquiry* **51(1)**: pp. 101–121.

Olper, A. & V. Raimondi (2009): "Patterns and determinants of international trade costs in the food industry." *Journal of Agricultural Economics* **60(2)**: pp. 273–297.

Owen, D. & N. Winchester (2014): "The impact of us fresh milk production standards on dairy trade." *Journal of Policy Modeling* **36(6)**: pp. 1008–1021.

Polák, P. (2017): "The productivity paradox: A meta-analysis." *Information Economics and Policy* **38**: pp. 38–54.

Polák, P. (2019): "The euro's trade effect: A meta-analysis." *Journal of Economic Surveys* **33**: pp. 101–124.

Pöyhönen, P. (1963): "A tentative model for the volume of trade between countries." *Weltwirtschaftliches Archiv* **90**: pp. 93–100.

Ravenstein, E. G. (1885): "The laws of migration." *Journal of the Statistical Society of London* **48(2)**: pp. 167–235.

Reed, W. R. (2015): "A monte carlo analysis of alternative meta-analysis estimators in the presence of publication bias." *Economics-The Open-Access, Open-Assessment E-Journal* **9**: pp. 1–40.

Robertson, R. & A. Estevadeordal (2009): "Gravity, bilateral agreements, and trade diversion in the americas." *Cuadernos de Economía* **46(133)**: pp. 3–31.

Romalis, J. (2007): "Nafta's and cusfta's impact on international trade." *The Review of Economics and Statistics* **89(3)**: pp. 416–435.

Sanjuan, A. I., M.-L. Rau, D. Oudendag, & M. Himics (2017): "Analysing eu dairy exports: indicators of non-tariff measures and gravity." *Technical report*.

Silva, J. S. & S. Tenreyro (2006): "The log of gravity." *The Review of Economics and statistics* **88(4)**: pp. 641–658.

de Sousa, J., T. Mayer, & S. Zignago (2012a): "Market access in global and regional trade." *Regional Science and Urban Economics* **42(6)**: pp. 1037 – 1052. Special issue 'In honor of Jacques Thisse'.

de Sousa, J., T. Mayer, & S. Zignago (2012b): "Market access in global and regional trade." *Regional Science and Urban Economics* **42(6)**: pp. 1037 – 1052.

Stanley, T. D. (2005): "Beyond publication bias." *Journal of economic surveys* **19(3)**: pp. 309–345.

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 (2017): "Neither fixed nor random: Weighted least squares meta-regression." *Research synthesis methods* **8(1)**: pp. 19–42.

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.

STEEL, M. F. (2019): "Model averaging and its use in economics." *Journal of Economic Literature* **forthcoming**.

STEPHENS, B. (2015): *America in retreat: The new isolationism and the coming global disorder*. Sentinel.

THARAKAN, P. K. M., I. V. BEVEREN, & T. V. OURTI (2005): "Determinants of india's software exports and goods exports." *The Review of Economics and Statistics* **87(4)**: pp. 776–780.

THURSBY, J. & M. THURSBY (1987): "Bilateral trade flows, the linder hypothesis, and exchange risk." *The Review of Economics and Statistics* **69(3)**: pp. 488–95.

TINBERGEN, J. J. (1962): "Shaping the world economy; suggestions for an international economic policy." .

WRIGHT, J. H. (2008): "Bayesian model averaging and exchange rate forecasts." *Journal of Econometrics* **146(2)**: pp. 329 – 341. Honoring the research contributions of Charles R. Nelson.

XU, X. (2000): "International trade and environmental regulation: Time series evidence and cross section test." *Environmental and Resource Economics* **17(3)**: pp. 233–257.

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, & M. Larch (2016): *An advanced guide to trade policy analysis: The structural gravity model*. World Trade Organization Geneva.

Zhang, D. & L. Nguyen (2018): "Tariff and us paper products trade." *Forest Science* **65(1)**: pp. 77–86.

4.A Appendix

Table 4.6: List of studies

Study	Head and Mayer (2014)	WP
Adam (2019)	No	Yes
Ahn *et al.* (2011)	Yes	No
Anderson & Marcouiller (2002)	Yes	No
Anderson *et al.* (2018)	No	No
Arkolakis *et al.* (2012)	Yes	No
Asche *et al.* (2018)	No	Yes
Baier & Bergstrand (2001)	Yes	No
Bas *et al.* (2017)	No	No
Békés *et al.* (2012)	No	Yes
von Below & Vézina (2016)	No	Yes
Bergstrand (1985)	Yes	No
Bergstrand (1989)	Yes	No
Berman (2009)	No	Yes
Berman *et al.* (2012)	Yes	No
Besedeš & Cole (2017)	No	Yes
Blonigen & Wilson (2010)	Yes	No
Caliendo & Parro (2014)	Yes	No
Cipollina *et al.* (2013)	No	No
Cipollina *et al.* (2016)	No	No
Costinot *et al.* (2011)	Yes	No
Coughlin & Novy (2012)	No	No
De (2010)	No	No
De Melo & Solleder (2018)	No	Yes
de Sousa *et al.* (2012a)	Yes	No
Dutt & Traca (2010)	Yes	No
Eaton & Kortum (2002)	Yes	No
Egger & Pfaffermayr (2003)	Yes	No
Egger & Průša (2016)	No	No
Emlinger *et al.* (2006)	No	Yes
Erkel-Rousse & Mirza (2002)	Yes	No
Estevadeordal *et al.* (2003)	Yes	No

Continued on next page

Table 4.6: List of studies (continued)

Study	Head and Mayer (2014)	WP
Fink *et al.* (2005)	Yes	No
Flach & Unger (2018)	No	Yes
Fontagné *et al.* (2005)	No	No
Francois & Woerz (2009)	Yes	No
Fugazza & Nicita (2011)	No	Yes
Fuller & Kennedy (2019)	No	No
Gervais *et al.* (2011)	No	Yes
Hayakawa (2014)	No	No
Hayakawa (2013)	No	No
Head & Ries (2001)	Yes	No
Heid *et al.* (2017)	No	No
Hertel *et al.* (2007)	Yes	No
Hou *et al.* (2017)	No	Yes
Hugot *et al.* (2016)	No	Yes
Hummels (1999)	Yes	Yes
Cheong *et al.* (2018)	No	No
Chevassus-Lozza *et al.* (2005)	No	No
Jacks *et al.* (2010)	No	No
Jakab *et al.* (2001)	Yes	No
Jaroensathapornkul (2017)	No	No
Kinzius *et al.* (2018)	No	No
Lai & Trefler (2002)	Yes	Yes
Lanati (2013)	No	Yes
Liapis (2011)	No	Yes
Martinez-Zarzoso & Nowak-Lehmann (2003)	Yes	No
Matyas *et al.* (2004)	Yes	No
de Sousa *et al.* (2012b)	No	No
Mazhikeyev & Edwards (2013)	No	Yes
Melo *et al.* (2012)	No	Yes
Keita (2016)	No	No
Nahuis (2004)	Yes	No
Novy (2006)	No	Yes
Novy (2013)	No	No

Continued on next page

Table 4.6: List of studies (continued)

Study	Head and Mayer (2014)	WP
Olper & Raimondi (2009)	No	No
Owen & Winchester (2014)	No	No
Robertson & Estevadeordal (2009)	No	No
Romalis (2007)	Yes	No
Sanjuan *et al.* (2017)	No	Yes
Tharakan *et al.* (2005)	Yes	No
Thursby & Thursby (1987)	Yes	No
Xu (2000)	Yes	No
Zhang & Nguyen (2018)	No	No

Notes: WP = working paper, Head and Mayer (2004) = studies included in that study

Chapter 5

Response to Referees

I am grateful to all the referees for their comments and useful suggestions in their referee reports and believe these helped us improve the dissertation and expand my horizons. I find it a pity that I did not receive such insightful questions during the publication process, especially for my first article. Based on the referees' suggestions, pre-defense discussion, and committee recommendation, I address most of the issues in the introduction. I prefer the suggested option not to change the text of the published papers significantly (Chapters 2 and 3) even if I would do some things differently now. Rather, I explain myself and address the remarks of referees in this chapter. I find such an approach more transparent. Many questions address the differences between chapters, which is caused mainly by learning more about meta-analysis. Chapter 1 addresses the key shifts in the methodology and approach towards the meta-analysis and hence my work. The main comments by the referees are typeset in roman; my response is in italics.

5.A PhDr. Marek Rusnák M.A., Ph.D.

Thank you for the opportunity to referee this dissertation. The three papers which constitute Peter's thesis represent excellent and state-of-the-art economic research with clear contribution to the existing literature. Overall, the thesis is well-executed and clearly written. Therefore, I recommend the thesis for defense without substantial changes.

Thank you for these kind words and I am pleased to get them from you.

I have no major comments. Having said that, I think the introduction could be improved along several dimensions. First, I really appreciated the discussion of limitations in the short sections in the first paper. I think a similar, more general discussion of limits of meta-analysis could complement the rationale for using the meta-analysis and could be interesting for the readers. For example, while the introduction briefly touches when discussing the sufficient number of studies for meta-analysis, but the discussions could go beyond the call for more replication studies. Second, I believe that the discussion of the origins of publication bias (page 3) could be toned down a bit. For example, it could be mentioned that some of the decisions of researchers are entirely rational and justified at micro level (e.g. changing specification when the estimated coefficient is clearly at odds with theory) but biases the aggregate mean of the published estimates available in the literature. Third, the potential of meta-analysis to provide decision makers with aggregate estimate of the effect under investigation based on best-practice is not highlighted enough in the introduction. While it is very interesting and useful to understand what drives the heterogeneity of estimates, it is equally valuable to have an ultimate one number summarizing our knowledge about the effect based on the whole literature. Given that the best practice effects are estimated in the individual papers of the dissertation, I think mentioning this in the introduction is only fair.

Thank you for the valuable comment, which I tried to incorporate in the introduction. I fully agree with these points, one of the limitation of the meta-

analysis being the lack of replication studies. I mentioned the need for replication studies not only in the previous version of the introduction at several places, but also in the leading quotation of the thesis. A call for replication also goes in hand with the need for transparency. The origins of publication bias are various, and finding the careful words is necessary. I agree that the choices are rational, and many researchers do not perform a sort of data-mining exercise to get exactly what they want. The last point about the usefulness of meta-analysis in policy discussion is a great point that I also supported with a few references.

It appears that the approach to treatment of outliers is not consistent throughout the thesis. Sometimes the outliers are thrown out based on arbitrarily selected limits (e.g. limit of 20 in Chapter 4), sometimes not outliers are excluded (Chapter 3). Were only outliers in the effect sizes investigated? In my experience, the estimates of standards errors could sometimes be extreme as well and could make a big difference in weighted regressions. Some additional explanations on these issues could be useful.

Thank you for noticing that. I admit that the outliers' treatment differs in the thesis as I learned new and more appropriate approaches. I consider the quantitative approach that is applied in the newest Chapter 4 as the most suitable for meta-analysis. Next to it, there was a mistake in the description of handling the outliers in Chapter 4 - the limit was used only for plotting charts, but they remained in the data sample (The wording was obviously changed accordingly). Since I consider this an important part of any meta-analysis, a closer description of my understanding of this issue is provided in the introduction as well.

In Chapter 2 the regression results in Table 2.5 show that around half of variables in regressions is not significant, keeping them in increases noise and estimates of standard errors, some rationale for not going to more parsimonious specification could be included, especially in the presence of ambiguous conclusions author mentions.

There are three reasons for that. Firstly, the purpose of the research was *to identify which variables are significant and to see which ones are not (and compare the results with the already existing research). Secondly, the variable selection was inspired by similar works in that area, and hence similar variables were kept in the regression. Thirdly, at that time, I was inspired by other meta-analytic works that did not eliminate the non-significant variables. The main reason being as unbiased as possible. Today, with more variables at hand, more quantitative approaches with the elimination of non-significant variables are also my preferred approach.*

5.B Prof. W. Robert Reed Ph.D.

Comments follow below. None of my comments are "major" in the sense of requiring major changes to the thesis. The few "major" comments I make apply to common practice in the economics meta-analysis literature in general. In many cases, there is not a clear "correct" way to do things. However, I think a little discussion about the respective issues would be helpful.

Thank you very much for the evaluation of my work and suggestions that I tried to reflect in the thesis. Below, I only selected those remarks which require my reaction as remarks to wording and typos were fully incorporated.

Regarding Chapter 2. I found the discussion associated with equations (2.5) through (2.10) to be confusing. ... I think it would be helpful if the author clarified his discussion of equations (2.5) through (2.10), taking into account the comments above if he feels they are correct.

Thank you for the question, as well as the accompanying specifications. It would be great if I received similar comments from the referees in the publication process. I fully agree that the discussion might be confusing if one attempts to seek the traditional models in it. The purpose of that paper was to provide the reader with an overview of the meta-analytic literature and the logic behind the final specification taken from **?**, *plus the notation well as structure, is directly taken from the mentioned papers. I tried to polish the structure and remove one step, which was only confusing to the reader. I believe that the final specification with the proper explanation is suitable. To connect it to the specifications you provided. In stata, the estimations were done using the mixed command; hence it should be the "Mundlak" model. As far as the notation, I had some mistakes in the notation of the variances.*

Regarding Chapter 2. I really liked Figure 2.1 and the directing of attention to the fact that estimation procedures can sometimes heavily weight a very few studies. In this case, two studies accounted for approximately 80 percent of the total weight. If there are heterogeneous true effects, this means that the true effects associated with those two studies get singled out as being more important than the effects from other studies, solely because these two studies had more observations. This is why I personally think it is good to also report RE estimates, even though this is not commonly done in the economics meta-analysis literature.

A well-constructed point, thank you. I personally was also surprised by these numbers, and this is one of the reasons I am very cautious when using the weighted approach nowadays. I truly understand the logic and the arguments that prof. Stanley uses when he explains the WLS methods. His work is the reason why I now can be presenting this dissertation, but the more I estimate I collected the more I am convinced that not only WLS methods should be used. The most precise estimates provide us with the "best" information by definition. But only if the precision is correctly estimated. As meta-analysts, we are not fully convinced that WLS methods should always be accompanied by multiple robustness checks.

(Regarding Chapter 2) The interpretation of the results is not straightforward because the dependent variable is the t-statistic." Actually, WLS estimates the same coefficients as the unweighted specification, and thus the coefficients can be interpreted as if one was interpreting a regression with $\hat{\beta}_{ij}$ as the dependent variable.

This point is well acknowledged, and I agree that it might confuse someone, especially with an overview of practices presented as equations 2.5 - 2.10. I would like to add more context and explanation. In general, weighted and unweighted specifications can be interpreted the same. I focused on the publication bias and t-statistic itself; hence the interpretation of the results-focused only on the sign and significance, not magnitude. My bad that I limited the added value of the paper.

The second issue is related to the incorrect handling of the dependent variable in the gravity equation, which is the logarithm of trade. Very often (i.e., in approximately one-quarter of the cases), the sum of imports and exports is taken before the logarithm." I am not a trade economist, so I didn't understand this. Surely the author does not mean that some studies calculate the dependent variable as ln(exports)+ln(imports), as opposed to ln(exports+imports). Perhaps a little clarification here would be useful.

Thank you for this comment. The base for the issue is one the "silver medal mistake" described by Baldwin and Taglioni (2007). As it might be surprising, it means exactly what is stated. In addition to that, the average of export and import is often used as a dependent variable, which means the same mistake. Fortunately, if the researcher used only uni-directional trade, this does not happen.

(Regarding Chapter 3) The text says, "Using econometric methods to analyse the funnel plot – FAT-PET – meta-regression analysis can be employed to identify the signal of a non-zero true effect should one exist in the research record." I have three comments to make regarding the FAT. First, strictly speaking, I believe the FAT can only legitimately be applied when publication bias is based on statistical significance, as opposed to "incorrect" sign. That being said, it is common practice in the literature to apply the FAT when publication bias is suspected to be due to bias against incorrect sign. I don't recommend you do anything differently, but maybe you could mention this in the text. My second point relates to recent statements by Stanley and Doucouliagos (2019) in which they state, "All meta-analysis methods fail to distinguish a genuine effect from the artefact of publication bias reliably under common conditions found in economics research" (page 15 of their paper, see link below). In other words, Stanley and Doucouliagos argue that the FAT is not very good at identifying publication bias. This echoes an earlier finding by Alinaghi and Reed (2018). Still everybody uses it, including myself. Again, I don't recommend that you do anything differently, but it would be good if you mentioned this, at least in a footnote. Finally, as you recognize later on, the SE variable is often correlated with other study characteristics. This is why you instrument for it in Table 3.3. But if one believes that, perhaps one should discuss the significance of the SE variable in the full meta-regression, since other variables are being held constant there. Is it still significant? Does your conclusion about publication bias remain?

Thank you for this comment and suggestion for improvement. The findings of Alinaghi and Reed (2018) are already pointed to in the last chapter (unfortunately, the paper was not known to me at the time I submitted my paper presented in Chapter 3 into the journal). The key message of Stanley and Doucaouliagos (2019) that you point out is quite striking, given the fact about what is done in meta-analysis. But as we also discussed multiple times, there is no other better way to address the issue; hence it is better to have at least something there.

(Regarding Chapter 4) It would be nice to have an expanded discussion of how tariffs are measured. Are they effective tariffs? Are they some measure of weighted, statutory tariffs? If effective, isn't there an endogeneity issue here? For example, in the extreme, exceptionally high tariffs could drive trade to 0, so that effective tariffs would equal 0.

Thank you for this comment and suggestion for improvement. The fact is, that the vast majority of the studies use ad-valorem tariffs (statutory ones) and their definitions. It is the easiest way to combine trade data with information about tariffs.

(Regarding Chapter 4) The author uses the BMA estimates to calculate "Best Practice" estimates. This implies that the BMA estimates are "better" than the OLS estimates. However, aren't the OLS estimates BLUE? Thus, for the purposes of prediction, shouldn't one use OLS estimates to produce the most efficient prediction?

Thank you for this comment. This is a fair point. The reason for using BMA estimates is rather in the structure of the research. BMA estimates are also based on the results of OLS estimation. We use the results of BMA for the key interpretation and OLS as a robustness check. To stay in that line of interpretation, BMA estimates are used for the calculation of the prediction. That is my perception. Especially as the "best practice" is also very dependent *on what is thought to be the best practice (e.g. as many data points as possible or including all countries and control variables, etc.).*

5.C Prof. Geoff Pugh, M.Sc., Ph.D.

: There is a notable – and pleasing – progression from Chapter 1 to Chapters 2 and 3. Not only is the quality of the meta-regression analysis higher, but the candidate appears to be more sure footed in the field of international macroeconomics than in the field of productivity analysis. If I were reviewing for a journal, I would recommend major revisions for Chapter 1 but only minor revisions (at most) for Chapters 2 and 3. Chapters 2 and 3 are excellent work and I have benefitted from reading them. I will certainly recommend them to my colleagues and graduate students. ... However, given that Chapters 1 and 2 have been published, I would not recommend that they be changed. Any issues here can be addressed in a revised Introduction. As for Chapter 3, I leave to the candidate's discretion as to which issues – if any – to address to increase chances of favorable reviewing upon submission to a good journal.

Thank you for your kind words and the suggestion and possibility to address your comments in the introduction. Overall, I agree with every detail and every imperfection you noticed. I tried to address the evolution of my work in Chapter 1. I personally prefer to make only small adjustments to the already published works since they can be found in that exact form on the Internet, and I think it is better to accompany it with additional comments and explanations. Any suggestions towards the last chapter, which is not yet published, would be a pity not to incorporate.

These two statements seem to be inconsistent. 1) Even for model averaging, one can specify priors (ex-ante limits and restrictions), which affect the outcome. 2) Model averaging ensures that the final model specification is not handpicked. "Specify" and "Handpicked" seem, at least to me, to be similar if not synonymous. This inconsistency informs my general impression that the decision to use BMA was never seriously in doubt. However, I would expect greater effort to be devoted to justifying this choice. I would expect both advantages and disadvantaged to be considered. For example, as far as I know, BMA practice does not incorporate diagnostic testing – e.g. the reporting and assessment of the Ramsey test for unmodelled non-linearity. In this case, averaging takes place over models that include an unknown number that are misspecified and whose results are to an unknown extent biased. I am not hostile to the use of BMA. Indeed, under the influence of the "Czech" School of MRA, I have used it myself. However, no good purpose is served by using it uncritically.

Thank you for this comment; I tried to add an extra explanation for it. There is still a difference and mainly in the process itself. Firstly, there are many ways in which priors are defined. One has the possibility of put the restrictions on variables but does not have to do that. Secondly, even if one uses e.g., sign restrictions on variables, the final specification is not one and only arbitrarily selected, but there will be many possible model specifications that fulfill such criteria. From that model space, only some are estimated and used for the averaging and final results. To sum up, when defining the priors, one has to reason for them, and still, the results will be robust since priors only restrict the model space and do not select one and only specification.

BMA was chosen over the previous approaches as the superior one with many advantages. BMA practice does can incorporate the diagnostic testing, but not every BMA package does so. I agree that in my text I put more emphasis on the positive reasons for method/approach selection and do not try to find every possible disadvantage.

"... the third paper does not use weighting by standard error as baseline methodology following the up-to-date works like Havranek et al. (2017)." Having told us the approach you do not use, please mention the "up-to-date" approach that you do use. There are certain inconsistencies between the chapters (noted below) that should be acknowledged and discussed in the Introduction. I would be happy with an explanation in terms of "learning". As we learn more, we change our ideas and corresponding practices. To inform the reader about the candidate's intellectual journey, especially the reasoning that led to different approaches in successive chapters, should be part of the function of the Introduction.

I would like to thank for these kind words and very thoughtful suggestion on how to explain myself. On my PhD research journey I made some mistakes as well as choices that seemed obvious at that time and I must admit, that there is a lack of criticism in my research. The reason is mainly that my focus is on application of already established methods, I do not see much of the point in repeating their weak parts - that does not mean to do the regressions no matter what.

Like BMA, I have nothing against multi-level modelling as a way of handling within-study heterogeneity. However, I would like some recognition of the disadvantages. In particular, like random effects estimation more generally, multi-level modelling rests on very strong statistical assumptions. For example, do you know the distribution of the "multiple random effects" in your model? I would like to know what attempts have been made to test for whether or not these assumptions hold in the data. And, if such diagnostic testing has not been undertaken, why not? Maybe because nobody else does? That may be an excuse, but it is scarcely good econometric practice. ... In general, I am recommending a more balanced approach to justifying the proposed methodology. Rather than treating BMA and multi-level modelling as almost self-evident choices, let us see the terms of debate and some acknowledgement that these methods may have disadvantages as well as advantages. It is always good practice to acknowledge the limitations as well as the benefits of a particular approach to empirical analysis.

Thank you for this comment; this is a fair point. It is true that while writing the papers, one tries to focus on what is new and take the methodology as granted and well tested by the previous research. The issues you mentioned with the assumptions, which I realized later on, were the key reason why I did not use it further. I adopted the methodology from similar works that were not open about the limitations of the methodology; hence, I only followed its application. I admit that only some diagnostic testing has been undertaken and *none reported. It might be an excuse that nobody else did that, but the truth is, that should have been done, and I learned from that experience that it is better to use methods where the assumptions are known, and I am not surprised later on.*

Referring to Eq.2.8, we read that: "The explained variable is the t-statistic and not the estimate of the effect size." Are you sure? The structure of your model develops Equations 2.5 and 2.6 (WLS). In Eq.2.6, the coefficients have the same interpretation as in Eq.2.5: i.e. alpha zero continues to measure publication bias (if any); and beta continues to measure the authentic empirical effect (if any) "beyond" publication bias. In effect, the coefficients switch places but preserve their initial interpretation. In Chapter 2, this point is acknowledged (p.68): "… the interpretation of the coefficients in equation 3.4 is the same …" So there appears to be some inconsistency between the chapters. ... Alternatively, the authentic effect size for studies with particular characteristics – say, a "best practice" specification – by setting the relevant moderators to one (and irrelevant ones to zero). Stanley and Doucouliagos (2012) discuss this in Ch.5; see, in particular, pp. 96-99. If so (if my reasoning is correct), then it is not the case – as asserted – that "we can only interpret the sign and significance". Stanley, T. and Doucouliagos, H. (2012). Meta-Regression Analysis in Economics and Business Research. London: Routledge. I note that a variant of the above procedure is adopted in both Chapter 2 (p.78) and in Chapter 3 (p.124 and 127). Why not in Chapter 1?

Thank you for this comment and valid point. The same point was made by Prof. Reed. and I described it above. I would like to add more context and explanation. In general, weighted and unweighted specifications can be interpreted the same. I focused on the publication bias and t-statistic itself; hence the interpretation of the results-focused only on the sign and significance, not magnitude. My bad that I limited the added value of the paper.

Regarding the "best-practice" approach, I did not know then, that it would be a nice idea to calculate it. I found out later on and hence chapter 3 and 4 *contain the computation of it.*

In the production function (Equations 2.1 etc), Q is output in physical terms. Yet your variable of interest is "the value of the ICT capital" (p.20). Hence, without any discussion or even acknowledgement, the production function has been recast into value terms. Of course, this is the typical procedure in productivity studies, because data on physical output is rarely obtainable and never widely comparable. However, researchers do need to explain precisely what is being estimated and why. In particular, when the theoretical and empirical forms of models are markedly different, readers should be informed of this and the implications assessed. Your "important coefficient" presumably measures the "value of the output elasticity of ICT capital"; not "the elasticity of ICT capital".

Thank you for this comment and again a valid point. It should be an output elasticity and I modified the proper part of the chapter so that it is clear. In my defense, the productivity literature skipped the "output" word a lot, probably to save space. From the equation, it is also clear, what is measured, but I agree that just from reading the abstract or the conclusion, one might draw a false conclusion.

Why do you include both "precision" and (the number of) "Observations" in your models? Both are generally treated as alternative measures/controls in MRA. Did you explore the likely collinearity between these two variables?

Thank you for this comment. The fact is that I did not use any variable elimination procedure at that time, and I did not want to make an arbitrary choice about which variable not to include. These two variables are correlated, but there is no collinearity. Some modification of the number of observation is usually an alternative measure of precision - I also use it as an instrument for that as well as other meta-analysts. But e.g. in the last chapter, the instrument is weak, which means that the correlation was low. To sum up, it is ok, to include both in one regression.

"Kohli & Devaraj (2003) conclude that a larger sample size leads to a higher ICT payoff ...” Why? It is at least plausible to argue the contrary: i.e. larger sample size should – ceteris paribus – increase the precision of estimates, thereby reducing the incentives that lead to positive section bias and thus lower estimated ICT payoff.

Thank you for this question, but I am afraid that I am not able to answer it. I paraphrased the conclusion of another study. It is possible to argue the opposite way if one assumes that there is a publication bias that causes an upward bias. And I fully agree with you and this way of reasoning - a larger sample means a better and more precise estimate. If the publication bias is present, the larger sample should lead to a smaller ICT payoff.

The headline conclusion is stated as follows: “the underlying effect is identified to be around 0.003 ...” This I find genuinely frustrating. What does this mean? Is this the representative output elasticity of ICT expenditure? Or what? Who knows? Be aware that some readers will read only the introduction and the conclusion of your article.

Thank you for this comment and I am sorry to cause a frustration. It is the output elasticity of ICT expenditure.

“... following recent trends in meta-analysis.” OK, but your supporting references seem to indicate a local rather than a global phenomenon. The rest of us are catching on slowly! “The preferred methodology for meta-regression analysis is the BMA ...” This is pure assertion. You have not established this. If true, why did you not use it in Chapter 1? Again, there is some unexplained inconsistency. “... and to work with a large number of regressors at the same time.” In my experience, the jump from days to months of computing time is made rather quickly. The point I am trying to make here is not that I object to BMA in MRA studies. (I have found it useful.) Rather, I am suggesting a rather more critical stance regarding your chosen methods, particularly as these have changed over time. The Introduction would be the place for such reflection. The comparison with the established WLS approach is most useful.

Thank you for this comment and valid point for the BMA. As I also stated *during the pre-defense, there has been an evolution in my work and not all three papers were written at the same time. I also understand the suitability of a critical stance regarding the methods used. My position is, that I consider published research to be valid, and if the methodology has merit, I implicitly do not question it. Maybe I should, but I think the reviewers did so as well and the author managed to convince them as well. Therefore I take established methods, put effort into understanding them, but I do not see the point in repeating pros/cons if I can refer to the previous works that already did that. My personal stance is that I rather use the place in the paper for the benefits of the analysis and key takeaways rather than the discussion of methodology.*

"Primary studies, from which we collect the estimates, use different methods to estimate trade elasticity with respect to trade costs or exporter 'competitiveness', such as wage, exchange rate, prices and productivity." Does this give a coherent effect size? Or do we have an "apples and oranges" problem? Some discussion of this point would be useful.

Thank you for this comment. Some part of the differences can be coded; some others would cause apples and oranges. The statement might be misleading as I understand. The purpose of that sentence was to introduce the variety that can be found in the primary research, but not that we have that mix in the dataset itself. The dataset itself contains only prices and exchange rates. Also, on p. 118 we split the sample by identifying variables to make sure we do not mix apples and oranges.

A nice feature of this paper is that the variables coded reflect the large literature on gravity modelling, for example with respect to the treatment of multilateral resistance and the treatment of zero flows in the trade matrix. However, I do have a couple of points to make in this regard. 1. The use of Poisson regression was not introduced primarily to deal with zero flows in the trade matrix. It is, according to its proponents, the preferred approach to estimating a gravity model in principle, given the nature of the error term. 2. I was surprised to see no mention of dynamic gravity specifications that capture the role of historical influences on trade patterns. Do none of the studies in this literature specify dynamic models? If so, that might be worth a comment. For references to dynamic gravity specifications more generally in the trade literature, see, for example: Gashi, P, Hisarciklilar, M & Pugh, G 2017, 'Kosovo–EU trade relations: a dynamic panel poisson approach' Applied Economics, vol. 49, no. 27, pp. 2642-2654. https://dx.doi.org/10.1080/00036846.2016.1245836

Thank you for this comment and kind words of appreciation, it took a long time to compose the table. Regarding 1, I understand it, but most of the researchers use it in that way. Regarding 2) that is a crucial point and we will look more deep into it. Hopefully we will find more studies and will be able to expand our sample.

The results of the test presented in Table 4.2 come from several specifications of Equation 4.7, and all are linear." How do you know that your linear specification is supported by the data? Standard econometric practice would be to report and interpret the Ramsey test.

Thank you for this comment. We probably did not used our words carefully. The linearity described there means, that all the test are testing linearity. We know, that there are non-linear tests for publication bias, but we did not incorporated them yet.

You conclude: "Due to the data heterogeneity, conditional estimates (based on the preferred combination of explanatory variables) of trade elasticity have such large confidence intervals that the results are not significantly different from zero. Therefore, we cannot make any conclusions based on these data aggregates."
As I indicate above, I think you have done a good job of capturing the heterogeneity in this literature. As such, therefore, heterogeneity should not be an obstacle to capturing an authentic empirical effect – beyond both publication bias and heterogeneity – should such an effect be present in the data. In this case, your conclusion should be that meta regression of the literature provides no evidence of a substantial trade effect from the trade costs investigated. You

seem to shrink from the implications of your own analysis. Better would be to trust your own analysis, point to any limitations in your study, and sketch a research agenda designed to check the robustness of your findings.
Some of the effects that we economists take for granted just do not seem to be there in the data (or are present to only a rather minor extent). It is part of the mission of MRA to uncover such uncomfortable conclusions.

Thank you for your words of support. I tried to find a better formulation for that conclusion, but I did not find any. The analysis is heterogeneity is self-standing and separated from the best-practice. But for the best-practice, the limitation is the wide confidence interval and given what we have available, any conclusions cannot be made. I also think that interpretation that using best-practice, we cannot see any trade cost elasticity effect would be a wrong conclusion. I was honestly disappointed to see it, but we are open about the results. The future agenda is hard to align for meta-analysis since we use all estimates and the heterogeneity is just huge and we do not have any further ideas, how to capture the rest of it that would help us lower the confidence intervals.