Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

The Effects of Structural Reforms in Europe: A Meta-Analysis

Author: **Bc. Elena Mizeráková**

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

Academic Year: **2018/2019**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, January 4, 2019

Signature

Acknowledgments

I would like to thank my supervisor, doc. PhDr. Tomáš Havránek, Ph.D., for his patient guidance and valuable advice he has provided throughout my work on the thesis. I would also like to thank Prof. Nauro Campos for introducing me to the topic and giving me the opportunity to work on it. Finally, I would like to express my gratitude to my family and friends for their unwavering support and encouragement during my university studies.

Abstract

In this thesis, we use modern meta-analytical methods to conduct a systematic review of the literature that estimates the effect of structural reforms on economic performance in European countries. We collect 889 estimates from 90 studies, which we then test for the presence of publication bias. Using the regression-based tests, we find evidence of small publication bias. In order to examine why reform effects vary and tackle the model uncertainty, we employ Bayesian model averaging. Our findings indicate that data choice and estimation methodology are the most important factors in explaining the effect heterogeneity. While we do not find any significant short-run effect, we confirm the positive long-run effect of the reforms. Furthermore, fiscal and institutional reforms are shown to be the most effective type of reform in achieving long-term growth, as opposed to labour market reforms that are proved the least beneficial for economic performance.

JEL Classification	C83, E60, O11
Keywords	meta-analysis, structural reforms, growth, publication bias

Author's e-mail	`elena.mizerakova@gmail.com`
Supervisor's e-mail	`tomas.havranek@fsv.cuni.cz`

Abstrakt

V tejto práci sú využité metódy modernej meta-analýzy s cieľom vytvoriť systematický prehľad literatúry, ktorá skúma efekt štrukturálnych reforiem na ekonomický rast v krajinách Európy. Analyzujeme 889 odhadov tohto vzťahu publikovaných v 90 štúdiách, ktoré potom testujeme na prítomnosť publikačnej selektivity. S použitím testov založených na regresnej analýze nachádzame stopy malej publikačnej selektivity. U krátkodobých efektov pozorujeme malú publikačnú selektivitu. Aby sme mohli skúmať, prečo sa jednotlivé efekty líšia a k ošetreniu modelovej neistoty, používame bayesovské modelové priemerovanie. Naše zistenia naznačujú, že najdôležitejšími faktormi, ktoré vysvetľujú rozdiely medzi odhadmi, sú voľba dát a metodológie odhadu. Aj keď nenachádzame štatisticky významné krátkodobé dopady, práca potvrdzuje kladný dlhodobý dopad reforiem. Fiškálne a inštitucionálne reformy sa navyše ukazujú ako najefektívnejší typ reforiem v dosahovaní dlhodobého rastu, na rozdiel od reforiem trhu práce, ktoré sa ukázali byť najmenej výhodné pre ekonomický rast.

Klasifikace JEL	C83, E60, O11
Klíčová slova	metaanalýza, štrukturálne reformy, hospodársky rast, publikačná selektivita

E-mail autora	`elena.mizerakova@gmail.com`
E-mail vedoucího práce	`tomas.havranek@fsv.cuni.cz`

Contents

List of Tables viii

List of Figures ix

Acronyms x

Thesis Proposal xi

1 Introduction 1

2 Studies on Structural Reforms 3

2.1 What Structural Reforms Are 3

2.2 Approaches to Measuring the Impact of Reforms 4

2.3 Overview of Results from Empirical Studies 5

2.4 Meta-Analyses on Reform-Growth Effect 7

3 The Reform Effects Dataset 10

3.1 Data Collection 10

3.2 Collected Data Characteristics 12

3.3 Data Description 17

4 Estimating the Average Reform Effect 22

4.1 Measuring the Effect Size 22

4.2 The Short Run and Long Run Reform Effects 23

5 Testing for Publication Bias 28

5.1 Publication Selection Bias: What Is It? 28

5.2 Detecting and Quantifying Publication Bias 30

5.2.1 Funnel Plots 30

5.2.2 Meta-Regression Test of Publication Selection 34

6 Why Reform Effects Vary **40**
6.1 Methodology . . . 40
6.2 Implementing Bayesian Model Averaging . . . 42
6.3 Results . . . 45
6.4 Best-Practice Estimates . . . 50

7 Conclusion **53**

Bibliography **66**

A List of Primary Studies **I**

B Collected Reform Effects by the Reform Type **III**

C Results of BMA **V**

D Diagnostics of BMA **VIII**

E Robustness Check **XI**

List of Tables

3.1 Description and Summary Statistics of Regression Variables . . 20

4.1 Estimating the Average Reform Effect 27

5.1 Test of Publication Bias - Short Run Effects 38
5.2 Test of Publication Bias - Long Run Effects 39

6.1 Estimating the Average Reform Effect 52
6.2 Estimating the Elasticity of Growth 52

A.1 List of Primary Studies I

C.1 Coefficient Estimates (Short Run) VI
C.2 Coefficient Estimates (Long Run) VII

D.1 Summary of BMA Estimation (Short Run) VIII
D.2 Summary of BMA Estimation (Long Run) VIII

E.1 Coefficient Estimates (Short Run) XIII
E.2 Coefficient Estimates (Long Run) XIV

List of Figures

4.1 Mean of the Collected Reform Effects Differs in the Short Run and Long Run . . . 24

5.1 Collected Reform Effects for the Long Run and Short Run . . . 31
5.2 Funnel Plots of Different Types of Reforms . . . 33

6.1 Bayesian Model Averaging, Model Inclusion (Short Run) . . . 47
6.2 Bayesian Model Averaging, Model Inclusion (Long Run) . . . 48

B.1 Collected Reform Effects for the Short Run by the Reform Type III
B.2 Collected Reform Effects for the Long Run by the Reform Type IV

C.1 Correlation Matrix of Selected Study Characteristics . . . V

D.1 BMA Diagnostic Plots (Short Run) . . . IX
D.2 BMA Diagnostic Plots (Long Run) . . . X

E.1 Bayesian Model Averaging, Model Inclusion (Short Run) . . . XI
E.2 Bayesian Model Averaging, Model Inclusion (Long Run) . . . XII

Acronyms

2SLS Two-Stage Least Squares

BMA Bayesian Model Averaging

CLI Cumulative Liberalization Index

EBRD European Bank for Reconstruction and Development

FAT Funnel Asymmetry Test

GMM Generalized Method of Moments

MCMC Markov Chain Monte Carlo

MRA Meta-Regression Analysis

OLS Ordinary Least Squares

PET Precision Effect Test

PIP Posterior Inclusion Probability

PMP Posterior Model Probability

TFP Total Factor Productivity

WLS Weighted Least Squares

Master's Thesis Proposal

Author	Bc. Elena Mizeráková
Supervisor	doc. PhDr. Tomáš Havránek, Ph.D.
Proposed topic	The Effects of Structural Reforms in Europe: A Meta-Analysis

Motivation Structural reforms represent a popular tool for governments aspiring to achieve faster economic growth. Market deregulation, trade, institutional and financial sector reforms, if appropriately applied, may improve factor productivity and bring macroeconomic benefits. For governments designing the optimal policy and aiming to boost their economic performance, the knowledge of the effect that particular reforms have on the economy is of the utmost importance. Hence, over the years this topic has received a lot of attention in the academic community.

Plethora of authors have focused on reforms adopted in various countries, however, often with results that differ in direction or size of the predicted reform effect (McQuinn and Whelan, 2015; Bouis and Duval, 2011). The solution offers a quantitative procedure called 'meta-analysis'. This approach enables to understand the causes of variation and combine results of multiple studies, thus giving the most objective outcome. Meta-analysis has been widely used in economics, with its application in labour economics, Card et al. (2010) use it to evaluate labour market policies; in resource economics, Havranek et al. (2012) and Sebri (2014) determine price and income elasticities of gasoline and water, respectively; Doucouliagos and Ulubasoglu (2008) focus on the impact of political democracy on economic growth.

The basis for this thesis form two papers, Babecky and Campos (2011) and Babecky and Havranek (2013). The first identifies the main factors that explain the variation in estimated reform effects, while the latter further improves the analysis by correcting for publication bias and extends the discussion to the magnitude of the reform effect. The focus of both of the studies lies on transition economies. On the other hand, on European non-transition countries there has not been done any meta-analysis yet, which serves as an objective for this thesis.

Hypotheses

Hypothesis #1: The literature estimating the effect of structural reforms on economic growth is affected by publication bias.

Hypothesis #2: Numerical estimates of the reform effect depend on methodology employed to estimate them.

Hypothesis #3: Average structural reform has a strong positive long run effect on economic performance.

Methodology The main aim of the thesis is to investigate the responses of economic growth in Europe to various types of structural reforms using empirical research studies. The principal tool for achieving this objective is meta-regression analysis as described by Stanley (2001).

The first step when conducting meta-analysis is searching for all studies relevant to the topic. For this purpose, I will use standard databases (f.e. RePEc, EconLit, SSRN, Google Scholar), identify primary studies and their references. The sample will be reduced only to studies that report estimates of coefficients along with their t-statistics, clearly state the details of estimation methodology, and cover European economies. Consequently, I will classify important characteristics of each study, and determine the summary statistic that will serve as a dependent variable in my regression.

The coefficient estimates are usually not directly comparable among the studies, therefore, they will be standardized and evaluated both in the short and the long run. It is common practice to use for such evaluation so called "funnel plot", that will detect possible heterogeneity and publication bias. The latter refers to a state when results consistent with a certain theory or with statistical significance are more likely to be published (Stanley, 2005), and can deform the average reform effect. For example, if political expectations encouraged researchers to prefer strong or positive results, the average effect would be skewed. Based on the analysis, I will test for the presence of publication bias, and potentially correct the estimates. The regression model that includes variables accounting for differences among the studies will allow to determine the main sources of heterogeneity (similar to Babecky & Campos, 2011) and improve the estimates of reform effect. Finally, I will discuss the magnitudes and compare the responses of GDP to various types of reforms.

Expected Contribution I will perform a comprehensive quantitative review of literature estimating the response of economic growth to various types of structural reforms adopted in non-transition countries in Europe. I intend to use the most recent primary studies, modern meta-analysis methods and contribute to the limited number of studies concerning the link between reforms and growth. The thesis aims to be unparalleled in terms of its target countries and the scope of the reforms. The resulting estimates will give an accurate view on the effect different reforms have on the economy, and could provide valuable advice that can be broadly used in government decision-making. Therefore, the thesis may have significant impact on the current policy debate.

Outline

1. Introduction and motivation: This section will introduce the issue of publication bias, literature and meta-analyses on the link between reforms and growth.
2. Studies on reforms and growth: I will explain the most common approaches how researchers measure the reform effect.
3. Data set: I will describe the data collection process.
4. Empirical part: I will explain modern meta-analysis methods, test the presence of publication bias formally and graphically, correct the estimates, and build the regression model.
5. Discussion: I will discuss my regression results, sources of variation, the magnitude of a reform effect, and differences among the various types of reforms.
6. Conclusion: I will summarize my findings and their policy implications.

Core bibliography

Babecky, J., Campos, N. F., 2011. Does reform work? An econometric survey of the reform-growth puzzle. Journal of Comparative Economics. 39(2), 140-158.

Babecky, J., Havranek, T., 2013. Structural Reforms and Economic Growth: A Meta-Analysis. Czech National Bank.

Bouis, R., Duval, R., 2011. Raising potential growth after the crisis: a quantitative assessment of the potential gains from various structural reforms in the OECD area and beyond (No. 835). OECD Publishing.

Card, D., Kluve, J., Weber, A., 2010. Active labour market policy evaluations: A meta-analysis. The economic journal. 120(548).

Doucouliagos, H., Ulubasoglu, M. A., 2008. Democracy and economic growth: a meta-analysis. American Journal of Political Science, 52(1), 61-83.

Havranek, T., Irsova, Z., Janda, K., 2012. Demand for gasoline is more price-inelastic than commonly thought. Energy Economics. 34(1), 201-207.

McQuinn, K., Whelan, K., 2015. Europe's long-term growth prospects: With and without structural reforms.

Sebri, M., 2014. A meta-analysis of residential water demand studies. Environment, development and sustainability. 16(3), 499-520.

Stanley, T.D., 2001. Wheat from Chaff: meta-analysis as quantitative literature review. J. Econ. Perspect. 15 (3), 131-150.

Stanley, T.D., 2005. Beyond publication bias. J. Econ. Surv. 19 (3), 309-345.

Author Supervisor

Chapter 1

Introduction

In 2017, European economies were finally growing robustly at a rate of 2.4% per year. This came after a disappointing period between 2011 and 2016 when the average real growth rate reached only 1.5% and the economies were in dire need for reforms that would stimulate the economic activity. Many governments at that time proposed and implemented fiscal policies, and now it seems that their effort has been fruitful. We see that economic conditions improved, yet countries still face important challenges of how to improve long-term growth and how to make it more inclusive. (OECD 2018) Governments can focus on this goal and target various sectors of the economy with structural reforms, however, these must be designed carefully and systematically. Coordination of reforms, emphasis on their proper implementation, and the most suitable targeting can maximize the performance impact of structural reforms. The puzzle economists and governments are trying to solve is which area of the economy to target so that they achieve the desired effect most effectively and at the lowest cost.

Even though the indisputable popularity of the topic may indicate differently, the effects of structural reforms are surprisingly difficult to extract and the literature often delivers results contradictory in multiple ways. Despite the fact that the studies seem to estimate the same effect, some show that reforms bring gains even in the short run, while others conclude that economies can suffer short-run losses. The effects of different reforms are another source of dispute. In these cases, modern meta-analytical methods have proven useful. A meta-analysis, first defined by Glass (1976), is a formal quantitative study designed to assess previous research of the particular topic. The literature on the topic may be extensive, complex, and often conflicting, nevertheless, meta-analysis provides a coherent and consolidated review, because it studies empirical results as though they were any other scientific phenomenon. (Stanley 2001).

The objective of this thesis is to give an accurate view on the effect different reforms have on the economic performance, settle controversies from conflicting studies and improve the precision of the reform-growth effect. We collect 889 estimates of the reform-growth effect from 90 studies published between 1996 and 2017 and apply statistical methods to identify potential publication bias, explain sources of heterogeneity between the primary estimates, and ultimately construct the best-practice synthetic estimates.

Although there are two meta-analyses (Babecky & Campos 2011; Babecky & Havranek 2014) that estimate the reform-growth effect, both of them include only studies on transition economies. In this thesis, we expand the focus of existing meta-analytical research geographically to all European countries and above that, we differentiate between different types of reforms and compare their effects. We believe that this is the first work that examines the reform-growth effect in this sense. The results of the thesis also provide valuable advice that can improve understanding of how reforms influence growth, and that can be broadly used in government decision-making and policy design.

The thesis is structured as follows: in Chapter 2 we discuss what structural reforms are and we provide a review of available literature on structural reforms in Europe. Above that, we briefly summarize different research and estimation strategies along with their results. Chapter 3 covers the data collection process - we describe how we collected relevant studies and what the inclusion criteria were, we present collected study characteristics in our dataset and we comment on their descriptive statistics.Chapter 4 explains how we measure the reform effects and presents their averages not accounted for publication bias that might plague the literature. We follow with a definition of publication bias and we test whether it is present in the literature using funnel plots and formal tests in Chapter 5. Chapter 6 focuses on explaining the heterogeneity between study estimates and estimating a synthetic reform-growth effect. In Conclusion, we summarize our findings and provide concluding remarks.

Chapter 2

Studies on Structural Reforms

2.1 What Structural Reforms Are

Structural reform is a term used relatively loosely as there is no strict definition of what constitutes it. European Commission (2018b) provides a useful guide by explaining:

> "Structural reforms tackle obstacles to the fundamental drivers of growth by liberalising labour, product and service markets, thereby encouraging job creation and investment and improving productivity. They are designed to boost an economy's competitiveness, growth potential and adjustment capacity."

In other words, we can say that structural reforms affect the supply side of the economy and provide incentives to increase the quality and the quantity of the input factors (capital and labour) and to improve the technology (total factor productivity). Consequently, they should spur structural change, productivity and output growth. (Fischer *et al.* 2018) Hence, any policy that is introduced with the aim of enhancing growth prospects we consider a reform. We are interested in all kinds of reforms and we do not set any boundaries on any specific target sector.

For some policies, such as an increase in tax burden, it is clear what the reform and also its measure is. In contrast, a launch of an educational programme to reduce skill mismatch is not easily quantifiable so how does one measure it? Surely, it is difficult to assess particular regulation separately and out of context, because it depends on other regulations already in place. International institutions and researchers, therefore, periodically translate the economic environment into quantitative indexes and examine changes in these indicators that correspond to the conducted reforms. Examples of such indexes include the Transition Score by the European Bank for Reconstruction and Development (EBRD) and the Investment Reform Index by OECD. As a result of reform, the index can be upgraded, for instance when the Slovak Republic adopted a new law in 2018 incentivising investment in economically disadvantaged regions that addressed regional disparities (EBRD 2018b).[1] It can be, however, downgraded, just as in 2010 when the Slovak Republic made changes to the pension system that have made the operating environment for pensions more uncertain. (EBRD 2010)

2.2 Approaches to Measuring the Impact of Reforms

The question whether structural reforms really succeed in stimulating growth is subject to the analyses of plethora of authors. In general, there are three types of studies that focus on structural reforms and their effects: (i) theoretical studies, (ii) simulations, and (iii) empirical studies.

Usually, it is the international institutions, both European and worldwide, that issue reports and assessments of the reforming effort of individual countries (see for example, European Commission 2018a; Heipertz & Ward-Warmedinger 2008). These reports comment on the general progress of reforms and use expert knowledge to interpret the development of macroeconomic indicators. A common factor associated with these reports and papers is that they do not quantify the effect policy has had and only observe changes in macroeconomic indicators without taking into account other influences that might have affected them. Even though this approach is indeed very instructive, it might not be specific enough for future policy design and might be to some extent subjective.

In contrast to theoretical assessments, simulations build a forward-looking model based on the macroeconomic theory in order to evaluate the reform effect. These studies focus especially on taxes and labour market reforms and use either a micro-simulation or a general equilibrium analysis to simulate the reform effect on macroeconomic variables, often unemployment and growth (see for instance, Bouis & Duval 2011; Eggertsson *et al.* 2014; Gomes *et al.* 2013). The model-based literature agrees on the positive long-run effect of structural

[1] Methodology behind the EBRD transition indicators changed in 2017 and the key qualities considered include whether the economies are well-governed, green, inclusive, resilient and integrated.

reforms, but the conclusions for the short run differ. According to some (Gomes *et al.* 2013; Hobza *et al.* 2010) it seems that the benefits take several years to materialise, while Cacciatore *et al.* (2012) shows that both product and labour market reforms stimulate economic growth already in the short run (although job protection reforms temporarily increase unemployment). The results in this section of the literature strongly depend on the model setting and its calibration, similarly as empirical studies depend on model specification and dataset. Above that, model-based studies might lack validity because they may not reflect the reality properly. (Chattoe-Brown 1998)

We are interested the most in empirical studies that derive knowledge from actual experience rather than purely from theory and are estimated on directly observed and measured phenomena. Results in studies that we reviewed vary greatly because they are affected by the choice of data, methodology or model specification, however, they should still reflect the true response of the economy.

2.3 Overview of Results from Empirical Studies

The reviewed empirical literature is diverse - economists estimate different regression equations on different types of data (cross-sectional or panel) and treat potential endogeneity differently. When we look at the dataset and the corresponding model specification used, several reviewed studies average annual data over a few years (for instance, Jaimovich & Rebelo 2017; Heybey & Murrell 1999), so as to have cross-sectional data. This allows them to investigate the effect of structural reforms by estimating the following regression model:

$$g_i = \alpha + \beta R_i + \sum_j \gamma_j X_{ji} + \epsilon_i, \tag{2.1}$$

where g_i is a measure of economic performance for country i, usually annual real GDP growth or productivity growth, R_i is a measure of a particular reform analyzed in the paper, X_{ji} are explanatory variables that include initial conditions and other controls, and ϵ_i represents the error term. Regression coefficient β denotes the long-run, i.e. cumulative, effect of the reform measure on economic performance conditional on controls X_j, because of the averaged data. Yet, if the studies estimate the equation for a specific year, β describes a short-run effect. Ultimately in every regression, β captures three effects - (i) a direct reform effect through improved allocation of production factors, (ii) an indirect reform effect, e.g. through a positive effect on export growth, and (iii) the effect of other excluded growth determinants to the extent in which they are correlated with the reform measure.

Alternatively, many studies (among others Bouis *et al.* 2012; Apolte 2011) use panel data to estimate growth regressions of the following form:

$$g_{it} = \alpha + \beta R_{it} + \sum_j \gamma_j X_{jit} + \epsilon_{it}, \tag{2.2}$$

where g_{it} is a measure of economic performance of country i in year t, R_{it} is a reform measure for country i in year t. X_{jit} are explanatory (control) variables that include initial conditions and structural features for country i in year t, and ϵ_{it} denotes the error term. Variables α, β and γ_j denote the regression coefficients.

In panel data specification (2.2), β measures the long-run effect. Some researchers (among others Abed & Davoodi 2002; Havrylyshyn & Van Rooden 2003), however, want to differentiate between the long-run and short-run effects and estimate one of the upgraded versions of (2.2):

$$g_{it} = \alpha + \beta R_{it} + \delta R_{it-1} + \sum_j \gamma_j X_{jit} + \epsilon_{it}, \tag{2.3}$$

$$g_{it} = \alpha + \beta (R_{it} - R_{it-1}) + \delta R_{it-1} + \sum_j \gamma_j X_{jit} + \epsilon_{it}, \tag{2.4}$$

where R_{it-1} denotes the reform measure for country i in the previous year. In specification (2.3), the short-run effect is measured by the coefficient δ, while the long-run effect is the sum of coefficients β and δ. In contrast, β captures the short-run and δ the long-run effect in specification (2.4).

There are two major methodological issues the economists consider - unobservable country heterogeneity and endogeneity of the reform measure. The heterogeneity is usually overcome by estimating the reform-growth regression with country dummies or using panel data methods, e.g. fixed effects. Similarly, many studies control for time dynamics as well and include the time fixed effects in their model. A serious problem arises if the reform variable is endogenous which it likely is because as Rodrik (2005) points out, the policies are not random but used systematically to achieve certain goals. If the model is estimated with Ordinary Least Squares (OLS), it might lead to biased reform effects through reverse causality. Researchers often address this problem and use Two-Stage Least Squares (2SLS) or Generalized Method of Moments (GMM) (see examples in Beck & Laeven 2006; Piculescu 2003).

The findings in the reviewed literature on this topic are rather mixed. In general, economists confirm that reforms are beneficial in the long run, however, the size of their impact is disputable just as is their short-run effect. Some (Christoffersen & Doyle 2000; Wolf *et al.* 1999; Staehr 2005) show that reforms are associated with weaker output initially, but stimulate higher growth with one- to two-year lag, moreover, the long-run gains seem to dominate the immediate losses. As opposed to these studies, others (Bouis & Duval 2011) do not find any significant aggregate losses in the short run, and even argue that reforms may deliver some benefits, especially in good times.

The results differ also with respect to the target sector of the reform. A few studies compare multiple types of reforms, and those (among others De Melo *et al.* 1997; Fidrmuc 2003; Falcetti *et al.* 2002) in particular focus on transition economies which went through liberalization, privatization, and institutional reforms during the last three decades. To illustrate the disagreement between studies, for instance, Lawson & Wang (2005) find that price liberalization, enterprise reform, competition policy are negatively associated with growth and only trade liberalization is positive, while Staehr (2005) shows that broad-based reforms and liberalization bring gains, and only market opening and financial liberalization without accompanying reforms bring losses. Other papers focus on one type of reform only - particularly popular are financial and fiscal reforms, then also reforms of services, institutions, and labour market.

2.4 Meta-Analyses on Reform-Growth Effect

To our knowledge, there exist two meta-analyses investigating the effect of structural reforms. The first one was conducted by Babecky & Campos (2011). The authors gather 43 econometric studies, generating 321 estimates of the reform effect on economic growth, and closely examine the sign and magnitude of their statistical significance. All selected studies were published between 1996 and 2006 and were focused on liberalization and stabilization of transition economies. This particular focus of authors stems from four reasons: (i) popularity of the transition countries among the econometricians; (ii) similar initial conditions, but different timing and nature of the reforms; (iii) use of similar measures of growth, and (iv) use of similar econometric specifications. Approximately 30% of the coefficients found are positive and significant, which means that the reform improved the economic performance of the country. Another third of the coefficients are negative and significant, and the rest is not statistically significant.

The authors find that the use of panel data methods and time coverage play an important role in explaining the variation of estimated effects. Having a lower number of observations, using country-specific dummies and addressing potential endogeneity, all increase the probability of yielding positive reform-growth relationship. The relationship between growth and reforms is likely to be changing in time with reforms having initial costs that are offset in subsequent years, therefore the inclusion of lagged reform values increases the probability of a negative reform effect. On the other hand, estimates averaging the effects over several years are found to be more likely positive and significant. The sensitivity check shows that the results depend on the measure of reform used.

Babecky & Campos (2011) also draw attention to a potential drawback of their econometric analysis - the tendency of academic journals to publish analyses with statistically significant results. Babecky & Havranek (2014) address this issue of publication bias in their analysis, and extend the analysis of Babecky & Campos (2011) by including newly published studies and by deepening of the quantitative methods used.

The dataset used in Babecky & Havranek (2014) consists of 60 studies published between 1996 and 2013 that together contain 537 empirical estimates of the effect of different types of structural reforms on growth in transition and post-transition economies. The reforms could be categorized as (i) internal market liberalization, (ii) external market liberalization, and (iii) private sector entry reforms. In comparison to Babecky & Campos (2011), the authors distinguish between the short-run (within a year) and long-run (cumulative) effects of the reforms on growth.

The magnitudes of estimates across studies are generally not directly comparable, therefore the estimates are transformed using the corresponding t-statistics included in the primary studies. The standardized measure used by Babecky & Havranek (2014) is then the partial correlation coefficient that is further used in testing for the presence of the publication bias. That is confirmed only in the short-run estimates. The partial correlation coefficients are accordingly corrected, and the authors find that on average the reform-growth relationship is positive and small, while in the short-run negative and strong.

To examine the causes of heterogeneity among the results, the authors build a meta-regression model. Explanatory variables capture the design of primary studies, namely the econometric method used, the type of reform index, whether lagged variables were used and the time dynamics were controlled for, the control variables in the model specification; and publication characteristics. Because of the high number of variables (32) and uncertainty which of them to include, Babecky & Havranek (2014) use Bayesian Model Averaging (BMA) to estimate the model.As the most important factors in explaining the heterogeneity among the results authors identify the use of lagged variables and the use of external liberalization reform index. Overall, the results of the analysis show that a standardized reform in a transition country with 4% growth rate would in the short-term decrease the growth rate by 0.4 percentage points, but in the long-term increase by 0.3 percentage points. Furthermore, the reforms aimed at external liberalization are shown to have a greater impact than other types of reforms with lower cost in the short-term and greater benefits in the long-term.

Chapter 3

The Reform Effects Dataset

3.1 Data Collection

Every meta-analysis begins with a construction of data set that consists of all available studies focused on the phenomenon of interest. These studies are referred to as primary econometric studies. At this point, we have to define clear inclusion criteria that the primary study has to satisfy to be included in the data set. It is crucial to balance the inclusion criteria - if they are set too broadly, the studies satisfying them might be too heterogeneous and the meta-analysis results might be difficult to apply to specific cases. On the other hand, if the inclusion criteria are too narrow, it might become hard to find suitable studies and to generalize the results.

We define three inclusion criteria that the studies must satisfy to be included in our dataset. First, the study must investigate the effect of a reform on a key macroeconomic performance variable such as economic growth and report econometric analysis. We do not constrain ourselves on any particular type of reform, as we specifically wish to explore the differences between them. Second, the study must consider reforms in European countries, because we want to ensure the greatest possible relevance of the resulting recommendations for the European macroeconomic environment. Third, the study must include a description of the methodology used and the results of an econometric analysis with a measure of statistical significance such as t-statistics, standard errors or p-values.

In order to find suitable studies, we turn our attention to research databases, namely SSRN, RePEc, Google Scholar, and EconLit. As a searching query we use keywords:

`reform`, `growth`, `Europe`.

For the returned searching results we go through their abstracts to determine the initial studies' relevance. Many studies are purely theoretical or otherwise unsuitable, therefore dropped from our analysis. Subsequently, we go through the text itself to identify traditional empirical studies that estimate the effect of reform on a performance measure. A large part of the reform-growth research focuses on a simulation-based analysis using endogenous growth models such as DSGE. These studies have to be excluded because they analyze a theoretical impact of a shock on the economy, not an ex-post material one. The studies estimating the effect of the reform on unemployment are another large category of studies that have to be excluded. [1] Even though we include studies estimating the effect on productivity growth, unemployment is not directly translated to output growth.

Additionally, we go through the references of the relevant studies and include those satisfying our criteria to the dataset as well. Finally, we add primary studies from Babecky & Havranek (2014) that estimate the effect of structural reforms in transition economies. There have been a few cases in which the same reform-growth estimates were published in multiple studies under a different title or publication year. We pay close attention to not having any study estimates duplicated as we would unintentionally assign a greater weight to them and consequently bias our analysis.

In almost all primary studies their authors estimate more than one model specification and generally, the number of models estimated varies from study to study. Since the authors rarely clearly specify which model specification is preferred by them, we have decided to use all reform effect estimates from each study. This approach yields an unbalanced dataset, therefore in the further analysis, we assign weights to all estimates so that each study has the same importance. There are three advantages to this approach. First, we avoid biasing the meta-analysis by subjective selection of the authors' preferred estimates. Second, we increase the variation in our dataset, because many studies estimate alternative specifications and robustness checks. Third, using multiple estimates per study, we can remove the unobserved individual study characteristics by employing study-level fixed effects model (Havranek & Irsova 2017).

The final dataset consists of 90 studies satisfying the inclusion criteria. We finished our search in June 2018. The complete list of studies is included in

[1] The majority of studies that estimate the effect of reform on unemployment focus on labour market reforms.

Appendix A. The dataset includes both published and unpublished studies that contain together 889 empirical estimates of the reform-growth effect in European countries.

All studies in our dataset were published in the English language. The five oldest studies we found dated back to 1996; the newest study was published in the first half of 2018. In comparison with the two previous meta-analyses (Babecky & Campos (2011); Babecky & Havranek (2014)), the studies in our dataset cover longer period, broader geographical area and variety of reforms.

The selected primary studies focus on a wide range of reforms, nevertheless, can be categorized into clusters based on the target sector of structural reforms:

Labour market This cluster includes reforms aimed at increasing labour utilization and market adaptability such as the introduction of strong employment protection legislation or reforms of labour market institutions; also includes using other financial incentives to work (f.e. changes in public wage bill spending during the financial crisis).

Public finance and institutions Broad category of reforms that focus on taxation, quality of public finances, and capital regulation. Examples of such reforms include a reduction of the tax burden, institutional harmonization, and fiscal consolidation.

Innovation and financial markets Cluster includes reforms stimulating productivity growth - promotion of FDI, efficiency improvements towards financial services, strengthened property rights, commercial integration with the EU, globalization, and development of service-exporting sectors.

Welfare state Welfare state reforms address mostly demographic issues such as ageing, health care, and education. In particular, these include pension reforms of the 2nd pillar, changes in health care expenditure, welfare spending and European welfare models.

3.2 Collected Data Characteristics

From each study, we extract coefficients on the reform variables, its t-statistics, and additional characteristics that should help us explain the heterogeneity among the estimates.[2] Compared to the meta-analysis by Babecky & Havranek (2014) we have decided to gather additional characteristics reflecting the model specifications of our broader dataset. The gathered variables capture the data and the methods used, the specification characteristics and the publication characteristics. Beside the coefficients and t-statistics we extract the following information:

year: the year when the study was published. In case of unpublished studies, the variable represents the year when the study first appeared in the databases, for example as a working paper.

ref_type: the reform type of the study. Four reform types are coded: (1) labour market reforms, (2) public finance and institution reforms, (3) innovation enhancing and financial market reforms, (4) welfare state reforms.

panel: dummy variable equal to one if the study uses panel data. An alternative to using panel data are cross-sectional analyses that are generally used to estimate long-run reform effects. Panel data can be used to estimate both short and long-run effects, depending on whether the estimation includes only contemporaneous or also lagged measure of reform.

endo: dummy variable equal to one if reform measure endogeneity is treated in the estimation. Dummy hence corresponds to using one of the following statistical techniques: instrumental variables, 2SLS, 3SLS, GMM or cointegration.

fixed: dummy variable equal to one if the country-specific heterogeneity in the model is treated using either fixed effects model or country dummies.

n: the number of observations used for the model estimation in the primary study.

k: the number of independent variables used in the estimation of the reform-growth effect; including the constant, time or country dummies.

start: the first year of the data sample which was used for the reform-growth estimation in the primary studies.

tspan: the number of years in the data sample which was used for the reform-growth estimation in the primary studies.

[2] In case only the standard error or p-value is stated, the corresponding t-statistics is calculated based on this significance measure.

countries: the number of countries in the sample used for the reform-growth estimation.

gdppc: dummy variable equal to one if the dependent variable is GDP per capita growth. Other economic performance measures used as a dependent variable include GDP levels and Total Factor Productivity (TFP) growth, that is directly translated into the output growth.

ebrd: dummy variable equal to one if the EBRD Transition indicators are used as a measure of reform. EBRD Transition indicators include three reform components - internal and price liberalization (*lii*), external liberalization (*lie*), and private sector entry (*lip*). (EBRD 2018a)

cli: dummy variable equal to one if the Cumulative Liberalization Index (CLI) is used as a measure of reform. CLI by the World Bank also includes three reform components - *lii*, *lie*, and *lip*. (De Melo *et al.* 1996)

comb: dummy variable equal to one if a combination of CLI and EBRD indices is used as a measure of reform.

other: dummy variable equal to one if other measure of reform than CLI or EBRD index is used. Examples of such measures include an increase in minimum retirement age, a decline in product market regulation and a decline in labour tax wedge.

lii: dummy variable equal to one if internal markets liberalization components are used as a reform measure. These components include price liberalization and the abolition of state monopolies.

lie: dummy variable equal to one if trade and foreign exchange components are used as a reform measure. These components include liberalization of foreign trade and currency convertibility.

lip: dummy variable equal to one if banking reform, privatization or private sector entry components are used as a reform measure. These components include small-scale privatization, large-scale privatization, and interest rate liberalization.

av: dummy variable equal to one if an average (weighted or simple) or a sum of reform index components is used as a measure of reform.

margeff: dummy variable for the marginal reform effects; equal to one if *lii*, *lie*, and *lip* are used in the same specification.

lagdp: dummy variable equal to one if the lagged depended variable is included in the specification.

speed: dummy variable equal to one if the speed of reform is used as a measure of reform.

lags: dummy variable equal to one if both contemporaneous and lagged reform measures are included in the regression.

time: dummy variable equal to one if the time dynamics is controlled for in the specification.

ic: dummy variable equal to one if the regression specification controls for initial conditions. These controls might include initial GDP per capita, natural resources, economic distortions inherited from the past, geographic circumstances or initial life expectancy.

ic12: dummy variable equal to one if the regression specification includes initial condition measures in form of principal components estimated as in De Melo *et al.* (1996). The variable is relevant for the studies estimating the effect of structural reforms in transition economies.

nic: the number of types of controls for initial conditions.

stabil: dummy variable equal to one if the regression specification controls for stabilization. These controls might include inflation, fiscal balance-to-GDP ratio or budget deficit.

nstabil: the number of controls for stabilization (including the inflation if controlled for).

infl: dummy variable equal to one if inflation is controlled for in the regression specification.

inst: dummy variable equal to one if institutional development is controlled for in the regression specification. The controls might include legal and political indicators, democracy measures, political instability and violence, rule of law or government effectiveness measures.

ninst: the number of types of controls for institutional development.

fact: dummy variable equal to one if factors of production are controlled for in the regression specification. These controls might include investment per capita, human capital endowment, investment-to-GDP ratio or population growth.

nfact: the number of types of controls for factors of production.

open: dummy variable equal to one if a measure of openness is included in the regression specification.

init_GDP: dummy variable equal to one if the initial GDP is included in the regression specification.

fdi: dummy variable equal to one if foreign direct investment is included in the regression specification.

school: dummy variable equal to one if a measure of school enrollment (primary, secondary or tertiary) is included in the regression specification.

empl: dummy variable equal to one if an employment rate is included in the regression specification.

service: dummy variable equal to one if a share of the service sector is included in the regression specification.

pubpr: dummy variable equal to one if the study differentiates between the effect of the reform on the private and the public sectors.

journal: dummy variable equal to one if the study was published in a refereed journal.

lgoogle_pa: the number of study citations on Google Scholar per year since the study was published, i.e. the number of total citations divided by the age of the study in logarithm.

lgoogle_total: the total number of study citations on Google Scholar since the study was published in logarithm.

authaff: dummy variable equal to one if all authors are from academia.

3.3 Data Description

Table 3.1 lists all the variables we collect from the primary studies along with their summary statistics. We code many characteristics as dummy variables, therefore, their mean values represent the proportion of the total number of observations for which the variable equals one. We further distinguish short-run and long-run estimates and categorize the collected variables into five groups. The first group of variables describes the dataset and methodology used. The second group differentiates between the measures of reform used in each study. The third group of variables captures how the authors of each study accounted for the dynamics. The fourth group of variables describes what variables the authors controlled for and the fifth group describes the publication characteristics.

Data and methods The mean short-run reform effect in our sample is estimated using dataset starting in 1989 with 22 countries and 13 years. 99% of the short-run estimates come from specifications using panel data which is related to the fact that short-run effects are normally estimated including both contemporaneous and lagged variables. On the other hand, 82% of long-run estimates in our sample come from studies using panel data, and the mean effect is estimated on data starting in 1989 consisting of 23 countries and 13 years. Only slightly more than a third of all studies treats potential endogeneity of the reform variable in their specifications. The same holds about treating potential country-specific heterogeneity - only 38% of short run and 33% of long-run effect estimates come from specifications using fixed effects or country dummies. The average number of explanatory variables in regressions estimating the short-run effect is 16, and for the long-run effect 12, which is also related to the way short-run effects are estimated, i.e. using additional lagged variables.

Type of reform measure Most of the studies that estimate the effects of structural reforms in transition economies use as a measure of reform EBRD transition indicators or CLI index. Among these, some studies differentiate between the individual reform components, but 46% uses an average of the components. The rest of the studies - approximately a third - uses a variety of other measures of reform, for example, European Index of Regional Institutional Integration (Comunale & Mongelli 2018), Fiscal Rule Strength Index (Gavriluţă *et al.* 2017) or the size of the public sector (Magazzino 2014).

Measure of dynamics As expected, the majority of short-run studies include both contemporary and lagged reform variables. This is because within the same model setting, the reform variable without any lag allows to capture the short-run effect and the lagged ones capture the long-run influence. In case of studies estimating the long run effect, the proportion of studies using also lagged reform variables is significantly smaller, 37%. Then, over one-fifth of studies accounts for time dynamics using time dummies or time trend which eliminates seasonal components from the time series and enables better manifestation of the important patterns. Approximately 20% of short run and 12% of long-run studies use the speed at which the reforms were introduced and implemented as a measure of reform. This is typical for studies on transition economies in particular, because the authors often derive the measure from EBRD transition scores, for instance using change in reforms (Heybey & Murrell 1999) or dividing the countries into radical, gradual and lagging reformers (Wolf *et al.* 1999). In addition, slightly less than 24% of short run and 14% of long-run studies included in their model specification also the lagged dependent variable, which in other words means that their authors expected that the current level of growth measure is determined by its past level.

Specification characteristics A vast majority of all studies measures the effect of reform on the GDP per capita growth; other dependent variables occurring in our dataset are TFP growth (Nicoletti & Scarpetta 2003; Griffith *et al.* 2010), TFP (Barone & Cingano 2011; Harrison 2011), and GDP levels (Fetahi-Vehapi *et al.* 2015; Kang *et al.* 2014). Regarding the explanatory variables, studies estimating the reform effect typically include four types of controls: initial conditions, stabilization, institutional development, and factors of production. Frequent control variables include in case of the long-run effects initial level of GDP, foreign direct investment and government expenditure. The specifications estimating the short-run effects often include openness and differentiate between the public and private sectors.

Publication characteristics Almost half of the estimates come from studies published in refereed journals and are conducted by authors from academia only. Furthermore, the mean number of Google Scholar citations suggests that studies included are highly-cited which further stresses the importance and relevance of the topic.

We collect 889 estimates from 90 studies, therefore, we can immediately see that some studies were more generous with their estimation than others. The number of estimates included in each study varies greatly - from 9 studies we take just one estimate, while from Havrylyshyn & Van Rooden (1998) only we take 57 estimates. If we do not account for this fact, studies with greater representation in our sample would have greater weight compared to the studies with a few estimates. Such a situation might bias especially our analysis of the study heterogeneity. In order to give all studies the same weight regardless of the number of estimates they provide, we weight all the collected estimates. Each estimate is weighted by the inverse of the total number of estimates from the study. The weight assigned to each estimate is then

$$w_{ij} = \frac{1}{n_i},$$

where w_{ij} is the weight for estimate j from study i, and n_i is the number of estimates from study i.

Table 3.1: Description and Summary Statistics of Regression Variables

	Short run		Long run	
Variable	Mean	Std. dev.	Mean	Std. dev.
Data and methods				
ref_type	2.263	0.629	2.305	0.625
n	623.254	4171.525	423.462	3176.045
k	16.386	12.500	12.268	11.160
panel	0.988	0.108	0.815	0.389
endo	0.395	0.490	0.387	0.488
fixed	0.375	0.485	0.333	0.472
start	1988.979	10.022	1989.536	8.730
tspan	12.590	9.689	12.929	8.977
countries	21.926	6.468	22.765	8.103
gdppc	0.923	0.267	0.820	0.385
Type of reform measure				
ebrd	0.339	0.474	0.384	0.487
comb	0.139	0.346	0.089	0.285
other	0.313	0.464	0.387	0.488
lii	0.071	0.257	0.042	0.200
lie	0.047	0.212	0.038	0.192
lip	0.112	0.316	0.093	0.290
av	0.460	0.499	0.440	0.497
cli	0.192	0.394	0.138	0.345
margeff	0.153	0.361	0.086	0.280
Measure of dynamics				
lagdep	0.236	0.425	0.135	0.342
speed	0.204	0.403	0.122	0.327
lags	0.552	0.498	0.371	0.483
time	0.251	0.434	0.180	0.385
Specification characteristics				
ic	0.625	0.485	0.451	0.498
ic12	0.209	0.408	0.142	0.349
nic	1.437	2.095	0.778	1.194
stabil	0.758	0.429	0.631	0.483
nstabil	1.271	1.065	0.909	0.905
infl	0.732	0.518	0.545	0.498
inst	0.327	0.470	0.425	0.495
ninst	0.454	0.746	0.751	1.094
fact	0.330	0.471	0.456	0.499
nfact	0.463	0.789	0.864	1.136
pubpr	0.118	0.323	0.044	0.204
init_GDP	0.053	0.225	0.104	0.306

Continued on the next page

Continued: Description and Summary Statistics of Regression Variables

	Short run		Long run	
Variable	Mean	Std. dev.	Mean	Std. dev.
Continued from previous page				
fdi	0.041	0.199	0.114	0.340
gov_exp	0.112	0.316	0.187	0.390
open	0.130	0.337	0.147	0.351
school	0.094	0.293	0.144	0.376
empl	0.000	0.000	0.013	0.113
service	0.021	0.142	0.022	0.147
Publication characteristics				
journal	0.389	0.488	0.436	0.496
lgoog_pa	1.670	1.741	1.819	1.561
lgoog_total	3.963	2.257	3.981	2.203
authaff	0.445	0.498	0.440	0.497

Notes: Statistics in the table represent non-weighted means. All variables except for Google Scholar citations are collected from the studies estimating the effect of structural reforms. The list of primary studies is available in Appendix A.

Chapter 4

Estimating the Average Reform Effect

4.1 Measuring the Effect Size

What makes meta-analysis the highest level of evidence and differentiates it from a narrative review, vote counting, and other research review methods is its ability to provide concise information on both the direction and magnitude of research findings. (Burns *et al.* 2011) As Shin (2017) explains, effect size is the key concept in meta-analysis and an essential part of the quantitative research reporting and hypothesis testing. It represents a quantitative index of research findings and is further used as a dependent variable in the meta-analytical process, in contrast to the study characteristics that are used as independent variables.

Generally speaking, the effect size measures can be categorized into unstandardized (e.g. regression coefficients) and standardized, that include three large families of the standardized effect sizes - d, r, and odds ratio. d-family can be used when comparing two means (e.g. pre- and post-treatment comparison), while odds ratio measures association among categorical variables (e.g. cohort studies in medical research). The most common effect size measures are from r-family, i.e. the correlation family. Those include various test statistics, such as the correlation coefficient and the coefficient of determination R^2. (Shin 2017)

The choice of the proper effect size is related to the research design in the primary studies. In some cases, the choice of effect size measure is straightforward - the regression coefficient, because some phenomena are commonly examined in a particular way based on their observed behaviour which makes coefficients easily comparable across the studies. For example, in case of the relationship between commodity demand and price many are interested in % changes, which is reflected in the popularity of the log-log regression model. This makes the estimates of different studies directly comparable because all of them then measure the elasticity of demand. Yet this is not applicable in case of a reform-growth relationship, and there is no single uniform measure of the effect. Therefore, in order to synthesize different regression models used to estimate the reform-growth relationship, we are required to transform collected estimates from primary studies to a common metric that represents the effect sizes (Aloe & Thompson 2013).

Looking at other meta-analyses, Babecky & Campos (2011) use t-statistics, because these convey the magnitude of the statistical significance. The problem with their approach is that the t-statistics depend on the number of degrees of freedom, hence are not directly comparable. Following another option employed in Babecky & Havranek (2014), we construct a standardized measure, that is a partial correlation coefficient, accounting for the differences in units and variable transformations. The partial correlation coefficient r corresponding to the reform-growth effect is computed in the following way:

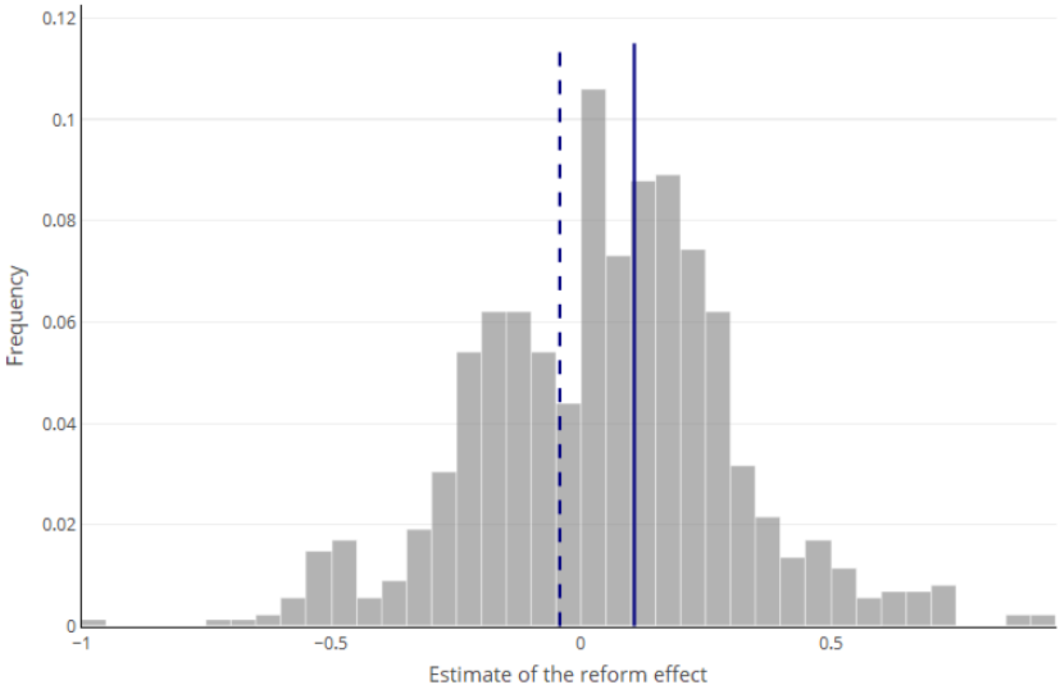
$$r = \frac{t}{\sqrt{t^2 + df}}, \tag{4.1}$$

where t denotes the t-statistic corresponding to the regression estimates found in the primary studies, and df denotes the number of degrees of freedom from the particular estimation. The range of partial correlation coefficient values is [-1;1]. The standard error associated with r is calculated according to the formula $SE(r) = r/t$ because the significance of the collected estimate must stay the same even after the transformation. The coefficient r itself allows us to assess the average effect of the reforms further in the analysis, and additionally, poses as a dependent variable in our meta-regression setting.

4.2 The Short Run and Long Run Reform Effects

All collected reform effects, i.e. the partial correlation coefficients, are illustrated in Figure 4.1. The figure depicts a histogram of 889 reform effects nor-

Figure 4.1: Mean of the Collected Reform Effects Differs in the Short Run and Long Run



Notes: The vertical axis shows the proportion of reform effects that fall into each size category; the horizontal axis shows the reform effect represented by the partial correlation coefficient. The solid line denotes the arithmetic mean of the estimates in the long run and the dashed line denotes the mean of the estimates in the short run.

malized to display relative frequencies of reform effect estimates. It shows the proportion of estimates that fall into each bin, with the sum of the bar heights equal to 1. Overall, the majority of the estimates is concentrated around zero and the histogram appears to be bimodal (or even multimodal). The mean of the short-run effects (dashed line) is negative in contrast to the mean of the long-run effects (solid line) that is positive. This suggests that the effects corresponding to the long run and short run may in fact differ and therefore, we first look at the short-run and long-run effects separately, and then further differentiate by the reform types.

Intuitively, negative mean for the short run and positive for the long run suggest that even though the reforms might have short-run costs, in the long run, they bring benefit. However, not all the reforms are created equal. In **??** we show the scatter plots of reform effects by the reform type. For the short run, the means of reform effect estimates are negative in case of labour market reforms, fiscal and institutional reforms, and welfare state reforms. On the other hand, financial and innovation spurring reforms seem to bring benefit even in the short run. For the long run, the reform effect means are positive for fiscal and institutional reforms, financial, and welfare state reforms. Only the labour market reforms estimates have a negative mean, thus are costly even in the long run. We further test this theory formally estimating the average effect by more specialized methods.

Three different measures of an average reform effect are reported in Table 4.1: simple arithmetic average, fixed effects estimator, and random effects estimator. All these estimates are first reported for all collected estimates and then for different types of reforms. What is the difference between the three average effect estimators?

If the reform effect estimates and their corresponding studies were all equally precise, the arithmetic average is entirely correct and is a simple estimate of the overall effect. Yet individual studies use different datasets, estimation techniques, etc.; all of that influencing their precision. We want to assign more weight to the more precise studies, and therefore we apply fixed effects or random effects estimators. In the fixed-effects analysis, the collected effect sizes are all assumed to be estimates of one true effect size. This means that if we could collect an infinite sample size for each study, the results would be identical across studies. The fixed effects estimator then weights the partial correlation coefficients using the inverse of their standard errors. On the other hand, in a random-effects analysis, the true effect is assumed to follow a particular distribution and we estimate the mean effect in this distribution. If we could again collect infinite samples for each study, the studies would result in different estimates. The random effects estimator weighs the partial correlation coefficients by the inverse of their standard errors and additionally takes into account the heterogeneity among the estimates. Because of the differences across studies, such as measures, methods, datasets, the random-effects method likely provides a better representation of real data. (Borenstein *et al.* 2010)

The findings established by the analysis of the histogram and scatter plots are confirmed empirically as well. The estimated arithmetic averages are -0.043 for the short run and 0.019 for the long run. Yet the estimates based on the more specialized meta-analysis techniques vary in terms of magnitude - the reform effect estimated by the random effects estimator is -0.041 for the short run and 0.106 for the long run. The fixed effects method predicts smaller reform effects with 0.042 for the long run and -0.009 for the short run. The direction of the effect is consistent across all three methods.

When we take a closer look at the different types of reforms, we see that the results are slightly more varied. For the long run, the average effects of fiscal, financial, and welfare state reforms are positive and significant. Based on the random effects estimator, the largest effect is predicted for the welfare state reforms, 0.105; the smallest effect is predicted for the fiscal and institutional reforms, 0.052. The labour market reforms are estimated to have a negative effect in the long run by the simple average, but this result is not significant; the other two measures even predict a significant positive effect. For the short run, the reform effects are negative in case of labour market and welfare state reforms - -0.043 for labour market reforms, and -0.038 for welfare state reforms. The estimated effects of financial and innovation spurring reforms are negative but not significant across all three methods.

Obviously, we would like to interpret the partial correlation coefficients and assess how important these effects actually are. For this purpose, we use the guidelines set by Doucouliagos (2011): the values of partial correlation coefficient smaller than 0.07 in absolute value denote no important effect; values between 0.07 and 0.17 denote a small effect; values between 0.17 and 0.33 denote a medium effect; values larger than 0.33 denote a strong effect.

The average reform effects we estimated suggest a small to near negligible effect on the economic growth in both the short and the long run. Still, the effects estimated in this section do not take into account the issues that may plague the literature and affect the effect sizes - publication bias and study heterogeneity. Publication bias is the reason why some estimates might have a different probability of being reported, while study heterogeneity reflects the differing models from which the collected estimates come from. Both these issues might skew the average reform effects, therefore we account for them in the following sections.

Table 4.1: Estimating the Average Reform Effect

	Short run			Long run		
Method	Est. effect	95% CI		Est. effect	95% CI	
All types of reforms						
Arithmetic average	-0.043	-0.070	-0.017	0.019	0.013	0.024
Fixed effects	-0.009	-0.013	-0.005	0.038	0.034	0.042
Random effects	-0.041	-0.067	-0.015	0.106	0.085	0.127
Labour market reforms						
Arithmetic average	-0.049	-0.108	0.011	-0.049	-0.119	0.021
Fixed effects	-0.010	-0.015	-0.004	0.023	0.020	0.026
Random effects	-0.000	-0.015	-0.014	-0.075	-0.135	-0.015
Fiscal reforms and institutions						
Arithmetic average	-0.095	-0.129	-0.060	0.136	0.110	0.162
Fixed effects	-0.083	-0.093	-0.073	0.127	0.119	0.134
Random effects	-0.094	-0.128	-0.060	0.138	0.114	0.162
Innovation and financial markets						
Arithmetic average	0.044	-0.003	0.090	0.073	0.030	0.116
Fixed effects	0.058	0.047	0.069	0.032	0.022	0.042
Random effects	-0.044	-0.002	0.091	0.067	0.031	0.103
Welfare state reforms						
Arithmetic average	-0.217	-0.768	0.334	0.014	-0.099	0.126
Fixed effects	-0.222	-0.438	-0.005	-0.019	-0.040	0.001
Random effects	-0.222	-0.438	-0.005	0.008	-0.098	0.114

Notes: Est. effect = estimated reform effect represented by the partial correlation coefficient; CI = confidence interval.

Chapter 5

Testing for Publication Bias

5.1 Publication Selection Bias: What Is It?

In the previous chapters, we described the evidence we collected from 90 studies that analyze the effect of reforms on growth. The topic of how the reforms affect macroeconomic outcomes is indeed very popular, and not only among researchers but also among politicians. It is the latter ones who may seek "the right" evidence in support of their reform proposal and may lead researchers to prefer reporting those results that are in line with the ideological view. Consequently, negative, non-confirmatory results and those that would contravene the politics may be systematically underrepresented in the literature and may cause distortions in the meta-analysis. Statistically speaking, many published scientific studies may then be a result of Type I errors (false positives) (Bradley & Gupta 1997) and many other studies may be left unpublished in the author's drawer. This problem, known as the "file drawer problem", was pointed out by Rosenthal (1979) and has been acknowledged by many authors in a variety of fields (e.g. in medical clinical research Hopewell *et al.* (2009); in social sciences Ashenfelter *et al.* (1999); Card & Krueger (1995)).

We have already mentioned that political influence or sponsorship can cause publication bias, however, what are the other possible causes? The journal editors, authors themselves and meta-analysts can all contribute to the publication selection more or less consciously. Beginning on the side of the journal editors, the first cause is related to their potential tendency to select more convenient results with respect to the theory, significance or expectation. For instance, the editors may believe that the studies with statistically significant results are more informative and therefore, they may prefer publishing them instead of the insignificant although likely still informative results. The same appears to hold also for studies consistent with the conventional view or theoretical knowledge - Card & Krueger (1995) point out that these might be published more likely than disagreeing studies.

In case of some of the phenomena investigated, the effect may be naturally constrained, e.g. the elasticity of a demand for some commodity. The elasticities are expected to be positive based on their definition, and thus the negative estimates may have a lower probability of getting published. Our reform-growth effect does not have such a constraint and may take on any value. Nevertheless, the publication bias induced by journal editors still might be present in our sample, because it is likely that mostly the results consistent with the major theory get published.

The authors themselves are another potential source of bias. They may choose not to submit their work because they might not consider it either important or interesting enough. According to Thornton & Lee (2000), theses and dissertations in psychological research are three times more likely to be published if they are positive and with statistical support for their hypotheses than if they are negative. Thornton & Lee (2000) believe that the main reason for this is the researchers' assumption that they will have a lower probability of being published, so they do not even try it. Another cause of bias on the side of the researcher is linked to the design of the study and its methodology. Studies using small samples usually have larger standard errors, therefore their results may lack statistical significance and be in turn less likely to be published. Some authors may then apply multiple models to their data and choose to report only the significant or otherwise important results, effectively skewing the evidence in the literature.

In addition, there are two occasions when the publication bias can arise during the execution of the meta-analysis as well. Acquiring unpublished studies requires a lot of effort as they are usually not accessible, and therefore there surely exist studies that are not included in the literature review. Furthermore, unpublished studies are not peer-reviewed, thus could be considered less reliable than the studies published in journals. Any flaws in unpublished studies included in the meta-analysis can consequently affect its results and bring in bias. Lastly, some studies relevant to the topic are not included in the review because of a language restriction. Many meta-analyses, this one included, collect papers in English only, however, there are possibly relevant papers also in other languages, which omission might have an effect on our final outcomes.

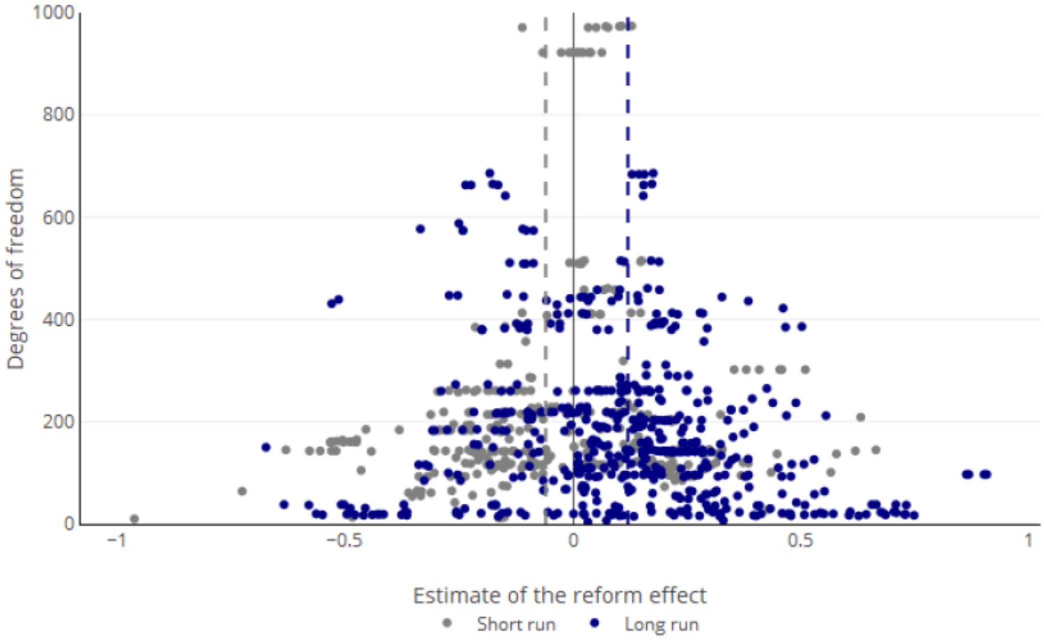
All the above-mentioned causes affect the published literature and might make the observed effect look more significant or strictly of a particular direction. Based on this aftereffect, we can distinguish two types of publication bias - directional selection and type II selection. Directional publication selection favours a particular direction of the observed effect which makes confirmatory results more likely to be published, and thus increases the number of studies with Type I error. With the second type, type II publication selection, statistically significant results with large t-values (in absolute value) are preferred regardless of the direction of the effect. This problem along with heterogeneity of true effect and misspecification biases result in excess variation among reported effects. (Stanley 2005) While type II selection itself is relatively benign and is unlikely to affect the conclusions of meta-analyses, directional selection brings serious distortions to the assessment of the true effect. The published evidence is skewed in the particular direction, shifting the mean effects even more into this direction and biasing the conclusion of the meta-analysis. (Stanley 2005) To avoid this scenario and to detect the publication bias, we apply two kinds of methods in the following sections - graphical methods (Funnel Graphs) and statistical regression methods.

5.2 Detecting and Quantifying Publication Bias

5.2.1 Funnel Plots

A funnel graph presents a simple visual representation of research literature and is the most commonly used method to test for the presence of publication bias (Sutton *et al.* 2000). The funnel graph is a scatter plot of a measure of precision on the vertical axis against the effect size (e.g. elasticity, regression coefficient, or in our case, partial correlation coefficient) on the horizontal axis. Precision can be measured as the inverse of the standard error, or alternatively as the sample size, its square root or the degrees of freedom employed in our analysis (Stanley 2005). Individual estimates should randomly vary around the true effect and due to heteroskedasticity form the inverted funnel shape from which the name is derived. Typically, small sample studies have also large standard errors (thus low precision measure) and form the broad part of the funnel, while with increases in sample size the precision should increase as well. Nonetheless, in the presence of publication bias, the randomness and symmetry are distorted. If one side of the graph is denser, the publication bias is likely

Figure 5.1: Collected Reform Effects for the Long Run and Short Run



Notes: The vertical axis shows the degrees of freedom used for estimation of the reform effect in the primary study; the horizontal axis shows the reform effect represented by the partial correlation coefficient. The solid line denotes zero; the dashed lines denote the arithmetic mean of the estimates in the short run (gray colour) and in the long run (blue colour).

present in that particular direction. If we suspect bias in favour of statistical significance, it is more difficult to infer from the graph. In that case the plot is wider and hollow, i.e. with greater variation, which does not automatically imply the presence of bias, but may be caused by other reasons. The most common are modelling misspecifications and temporal or spatial heterogeneity of true effects.

All 889 of our collected reform effects are illustrated in Figure 5.1, which separates the short-run and the long-run estimates. The short-run estimates are mostly scattered to the left from zero, while the long-run estimates to the right, but the funnels seem relatively symmetrical.[1] In case of the short run, the funnel has one peak around zero. At a closer inspection, some hints of publication bias appear, as the estimates are more concentrated on the left. On the other hand, the long-run funnel spreads widely across the x-axis indicating potential heterogeneity of the reform effect. The funnel is skewed to the left

[1] A different view with separate funnels for the short run and long run effects distinguishing the reform types can be found in Appendix B.

which again points to a slight publication selection.

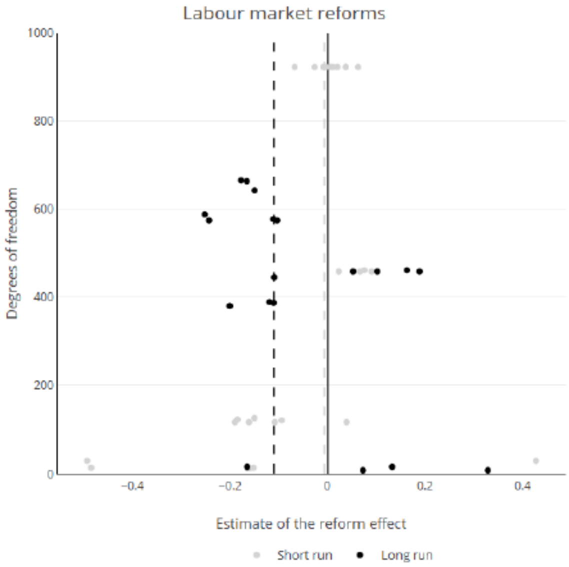
Figure 5.2 presents four funnel plots where the reform effect estimates are separated according to the reform type. On top of that, we separate the short run and long run effects due to the potential difference between these two as indicated before. Most of the observations belong to the fiscal and financial market and innovation reforms, which allows us to see the funnel shape more clearly in graphs (b) and (c), compared to the small number of observations for the labour market and welfare state reforms in graphs (a) and (d). In order to stress the difference between the short-run and long-run effects, we add a line signifying the arithmetic mean of observations in respective categories.

First, we focus on the labour market reforms in the short run. Figure 5.2(a) shows signs of a funnel forming with the most precise reform effects distributed around zero. The long-run effects vary significantly and are scattered over the chart. Because of the low number of observations, it is impossible to infer from the graph whether the publication selection is present or not. In case of welfare state reforms, we again see that the estimates are spread along the x-axis and differ widely, which suggests heterogeneity within the examined effect.

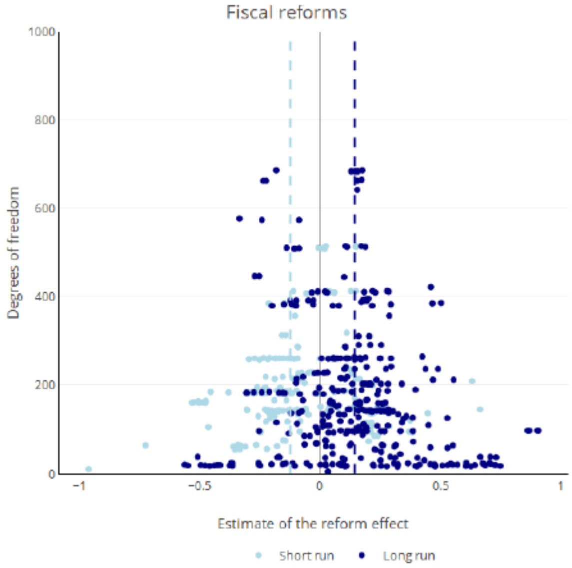
Figures 5.2(b) and (d) tell a different story and show a clear funnel. In case of fiscal reforms, the funnels for both the short run and long run effects do not appear hollow, which indicates that the type II selection might not be present. However, the short run funnel seems to be slightly skewed right and the long run funnel left, which suggests that publication bias might be present. The estimates of financial market and innovation reform effects vary greatly, as documented by the wide lower part of the funnel, but seem otherwise symmetric.

We find evidence of small publication bias, for fiscal reforms in particular, nevertheless, this method of funnel plot analysis is very subjective and our conjectures need to be examined more precisely with regression tests in order to draw an objective conclusion. Another limitation of the funnel graph is its implicit assumption that there exists a single "true" effect common to all studies. Heterogeneity of the reform effect due to the use of different data sets or different countries then might be a significant source of funnel asymmetry. On top of that, any misspecification in primary studies - omitted variables, inappropriate function form or estimation technique - may also induce funnel's skewness, thus we proceed with more advanced tests.

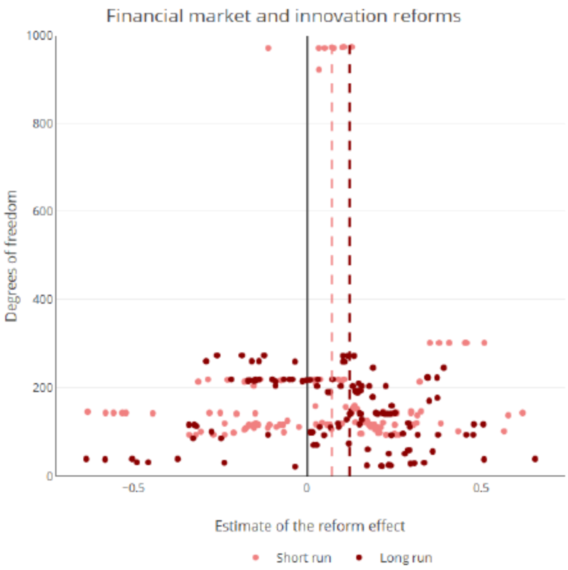
Figure 5.2: Funnel Plots of Different Types of Reforms



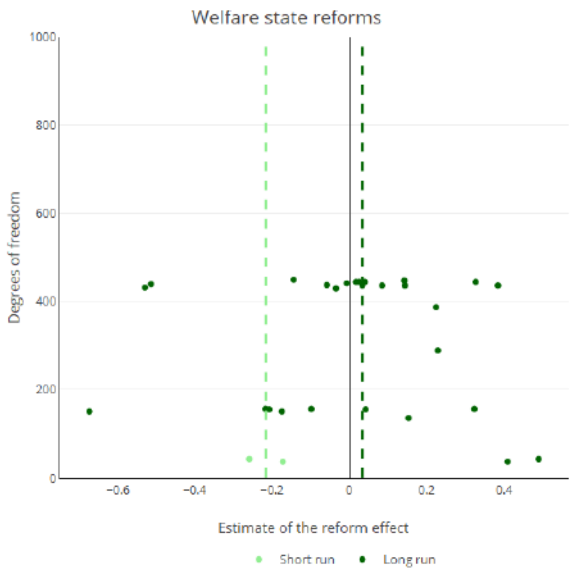
(a) Labour market reforms



(b) Fiscal and institutional reforms



(c) Innovation and financial market reforms



(d) Welfare state reforms

Notes: The vertical axis shows the degrees of freedom used for estimation of the reform effect; the horizontal axis shows the reform effect represented by the partial correlation coefficient. In all figures, the lighter shade denotes short-run effects, while the darker shade denotes the long run effects. The solid line denotes zero; the dashed lines denote the simple mean of the estimates in the short run and in the long run.

5.2.2 Meta-Regression Test of Publication Selection

We illustrated before that graphs are very intuitive and illustrative, yet they are also prone to the subjectivity of interpretation and a formal objective test of publication bias is required. The solution to this problem first proposed Card & Krueger (1995) in the form of Meta-Regression Analysis (MRA). Of course, when one thinks of MRA, the first thought might be that it is used to study the variation among reported econometric results. While this is the main purpose, it still allows for more. MRA can be used to model publication selection and statistical power, and to circumvent its effects (Stanley 2005).

For modelling publication bias, we use a simple MRA model that captures the relationship between the effects from primary studies and their standard errors as suggested by Stanley (2005). The regression equation has the following form:

$$r_i = r_0 + \beta_0 \times SE_i + u_i, \tag{5.1}$$

where r_i denotes the partial correlation coefficient from the i-th estimation, r_0 is the true value, i.e. the partial correlation coefficient corrected for publication bias. SE_i is the corresponding standard error for r_i, β_0 is coefficient measuring the magnitude and direction of publication bias, and u_i is the error term. In an ideal case, observed partial correlation coefficients (effects from primary studies) should vary randomly around the true value r_0 and be independent of their standard error (Stanley 2005). In other words, coefficient β should be statistically not different from zero. From the equation follows that if the publication bias is present and significant, it will be proportional to the standard error SE_i.

The problem with this MRA model is that the error terms u_i are likely to be heteroskedastic when using the OLS estimation. Each primary study uses different sample sizes and modelling specifications which makes the size of each error u_i likely to vary with studies, and as a result, the variance of all u_i is not constant ($Var[u_i|SE_i] = \sigma_i^2$). This violates the OLS assumption of independent and identically distributed errors u_i and can lead to misleading standard errors, and consequently narrow confidence intervals, t-statistics, and p-values. Hence, we apply four econometric methods that mitigate the heteroskedasticity and in turn the inefficiency when estimating the equation (5.1).

Within the first method, we correct the heteroskedasticity using standard errors clustered at the study level. Our key assumption here is that the error terms u_i are uncorrelated across studies while errors for reform effects belonging to the same study (i.e. cluster) may be correlated. Written in the language of statistics, we still assume zero conditional mean of our error terms u_{ij} corresponding to the reform effect i from study j, $E[u_{ij}|x_{ij}] = 0$, but we allow more flexibility in variance-covariance matrix. Therefore with clustering, $E[u_{hj}u_{ij'}|x_{hj}, x_{ij'}] = 0$, unless $j = j'$, where x_{ij} is a vector of independent variables, in our case just SE_{ij}. Clustering at the study level leads to more conservative standard errors (reduces the precision of $\hat{\beta}$), and thus we also employ a more liberal significance level ($\alpha = 0.1$).

Second, we employ estimation by Weighted Least Squares (WLS) as a suitable method for obtaining efficient estimates with corrected standard errors. With WLS, we first have to specify a model for $Var(r|SE)$, estimate it, and apply OLS to the observations weighted by an estimate of the conditional standard deviation $[Var(r_i|SE_i)]^{1/2}$. The weights are assigned to individual observations in such a way that they are smaller where the error term variance is large because the data is farther from the true regression. On the contrary, the weights are larger where error term variance is small because it contains relatively more information. An appropriate choice of weights that reflect the heteroskedasticity then leads to an efficient estimator, however, what weights should we choose? By an interesting twist, a sample estimate of the standard deviation of the dependent variable is our independent variable, therefore we can use the inverse of SE_i as weights. The MRA equation 5.1 then has the following form:

$$\frac{r_i}{SE_i} = r_0 \times \frac{1}{SE_i} + \beta_0 + \epsilon_i, \tag{5.2}$$

where $\epsilon_i = u_i \frac{1}{SE_i} \sim N(0, \sigma^2)$. Compared to the initial equation (5.1), the coefficients are reversed, but the coefficient β_0 still denotes the publication bias, while r_0 provides an estimate of the true effect size. To correct any remaining heteroskedasticity that the weighting has not eliminated, we estimate the equation (5.2) with clustered standard errors.

Another method that corrects for error-in-variable bias is instrumental variable estimation. It allows us to avoid biased coefficients caused by the fact that standard errors SE_i are estimates themselves and are correlated with error terms u_i. An appropriate instrument is the sample size n_i used for estimation of each reform effect r_i because it should be highly correlated with SE_i yet uncorrelated with SE_i's estimation error and thus exogenous to the errors u_i

(Stanley 2005). We estimate the equation:

$$r_i = r_0 + \beta_0 \times \sqrt{n_i} + u_i, \qquad (5.3)$$

where n is the sample size in estimation of r_i. Coefficient r_0 still stands for the true reform effect and β_0 quantifies the selection bias. On top of using an instrument, we again control for the heteroskedasticity with clustered standard errors when estimating (5.3).

Our fourth way of dealing with heteroskedasticity involves using the fixed effects, therefore we add a second random term to our MRA model. It captures the characteristics that are constant within the estimates from the same study yet are allowed to vary across studies. Our model then takes the following form:

$$r_{ij} = r_0 + \beta_0 \times SE_{ij} + \nu_j + u_{ij}, \qquad (5.4)$$

where ν_j is a vector of study-specific effects, assumed to be independently distributed as $N(0, \tau^2)$ and independent of both SE_{ij} and u_{ij}. Inherently, the fixed effects model assumes the existence of one true effect size and that all the studies estimate this effect on the same population using the same variables, etc. As a result, the true reform effect is a weighted average of study-specific effect sizes with more precise estimates contributing more to the summary effect than the imprecise ones (Borenstein *et al.* 2010).

In addition to these estimation procedures, we weight our observations by the number of estimates from each study as described in Chapter 3 to control for the fact we work with an unbalanced dataset. This approach is applied to all estimated models except for WLS, where we weigh by precision only.

Table 5.1 and 5.2 present the results of each test, separately for the short run and long run, and each reform type. For each set of results, we can comment on two aspects of interest - the publication bias and the effect corrected for it - basically testing two pairs of hypotheses. The first one, formally called Funnel Asymmetry Test (FAT), tests the null hypothesis $H_0 : \beta_0 = 0$, the publication bias is not present, against its alternative $H_1 : \beta_0 \neq 0$. The second one, Precision Effect Test (PET), then tests $H_2 : r_0 = 0$ against $H_3 : r_0 \neq 0$, i.e. investigates the mean value of reform-growth estimates after corrections for publication bias.

According to three out of four tests, publication bias is not statistically significant for the estimates of the short-run effects; only OLS indicates the presence of negative publication bias. PET does not find substantial evidence of a short-run effect as the corrected estimates have very low significance and differing directions. In contrast, the long-run effect is positive and very significant - reaches 0.15, which is larger than the averages estimated in Chapter 4 and it could be classified as a small to medium effect according to guidelines by Doucouliagos (2011). As regards the publication bias, FAT shows that the coefficient β_0 is significant in WLS and IV models, however, their results are contradictory in its direction.

When looking at the individual reform types, the test results vary widely. Regarding the short-run estimates of labour market effects, two of the tests confirm significant negative publication bias, however, the corrected reform effects differ greatly in their significance and direction, therefore we conclude that there is no substantial short-run effect. For the long run estimates, the story is very similar - we find some hints of positive selection bias and small negative reform effects, but only with fixed effects model.

In case of fiscal and institutional reforms, we do not find any evidence that the literature would be plagued by the publication selection. While there seems to be no short-run effect of this type of reforms on economic growth, in the long run, they have a significant positive impact reaching 0.22. Publication bias is neither confirmed for innovation reforms that in both short and long run show small positive effect that is mostly insignificant. We cannot assess the short run estimates of welfare state reforms due to lack of observations, yet for the long run, we find some positive publication bias and a positive corrected reform effect, although both are almost insignificant.

Overall, after correction for publication bias, we find that the reforms, especially the fiscal and institutional ones, bring substantial long-run benefit. As for the short-run effects, we do not find enough evidence for our hypothesis that reforms bring costs in the short run.

Table 5.1: Test of Publication Bias - Short Run Effects

	Short run			
	OLS	WLS	IV	FE
All reform types				
Publication bias (Coefficient β_0)	-1.037**	-0.595	0.001	1.531
	(0.429)	(0.490)	(0.001)	(3.163)
Effect beyond bias (r_0)	0.066*	0.007	-0.018	-0.122
	(0.040)	(0.010)	(0.043)	(0.232)
Labour market reforms				
Publication bias (Coefficient β_0)	-0.797***	-0.24	0.000	-1.032***
	(-0.084)	(0.505)	(0.000)	(0.023)
Effect beyond bias (r_0)	0.017	0.000	-0.054	0.031***
	(0.020)	(0.004)	(0.055)	(0.001)
Fiscal reforms and institutions				
Publication bias (Coefficient β_0)	-1.610	-1.834	0.023	5.354
	(2.040)	(1.777)	(0.016)	(5.809)
Effect beyond bias (r_0)	0.062	0.040	-0.389	-0.454
	(0.129)	(0.119)	(0.271)	(0.430)
Innovation and financial markets				
Publication bias (Coefficient β_0)	-0.936	-0.513	0.000	-1.485
	(1.254)	(0.510)	(0.001)	(11.110)
Effect beyond bias (r_0)	0.133	0.086*	0.057	0.173
	(0.113)	(0.041)	(0.058)	(0.810)
Welfare state reforms				
Publication bias (Coefficient β_0)	-	-	-	-
	-	-	-	-
Effect beyond bias (r_0)	-	-	-	-
	-	-	-	-

Notes: OLS - weighted by the number of observations for each study; WLS - weighted by precision; IV - sample size used as an instrument for standard error and weighted by the number of observations for each study; FE - with study fixed effects and weighted by the number of observations for each study. All regressions are estimated with clustered standard errors.

Table 5.2: Test of Publication Bias - Long Run Effects

	Long run			
	OLS	WLS	IV	FE
All reform types				
Publication bias (Coefficient β_0)	-0.080	1.350***	-0.002*	-0.878
	(0.305)	(0.266)	(0.001)	(0.802)
Effect beyond bias (r_0)	0.152***	-0.008	0.181***	0.266***
	(0.034)	(0.010)	(0.032)	(0.096)
Labour market reforms				
Publication bias (Coefficient β_0)	0.420	-1.671	0.000	1.722***
	(0.208)	(2.011)	(0.000)	(0.010)
Effect beyond bias (r_0)	-0.026	0.003	0.023	-0.145***
	(0.081)	(0.008)	(0.094)	(0.001)
Fiscal reforms and institutions				
Publication bias (Coefficient β_0)	-0.187	0.478	-0.007	-1.199
	(0.294)	(0.882)	(0.006)	(0.934)
Effect beyond bias (r_0)	0.221***	0.096	0.272***	0.354***
	(0.037)	(0.088)	(0.088)	(0.123)
Innovation and financial markets				
Publication bias (Coefficient β_0)	0.250	1.147	-0.003	1.922
	(0.842)	(0.696)	(0.002)	(1.669)
Effect beyond bias (r_0)	0.062	-0.024	0.126*	-0.099
	(0.073)	(0.022)	(0.063)	(0.160)
Welfare state reforms				
Publication bias (Coefficient β_0)	3.652*	2.812	-0.023	-1.041
	(0.996)	(1.866)	(0.008)	(0.900)
Effect beyond bias (r_0)	-0.103	-0.165	0.165	0.246*
	(0.162)	(0.085)	(0.417)	(0.067)

Notes: OLS - weighted by the number of observations for each study; WLS - weighted by precision; IV - sample size used as an instrument for standard error and weighted by the number of observations for each study; FE - with study fixed effects and weighted by the number of observations for each study. All regressions are estimated with clustered standard errors.

Chapter 6

Why Reform Effects Vary

6.1 Methodology

In the previous chapters, we have seen that the reform effect estimates vary widely across studies. Now, we wish to explain these differences and attribute them to individual study design characteristics. In order to do so, two commonly used approaches are available: Bayesian model averaging and frequentist meta-analytical method. We apply Bayesian model averaging as a preferred approach for several reasons.

To address the heterogeneity of the reform effect, we wish to build an expanded regression model from Chapter 5:

$$r_i = r_0 + \sum_j \beta_j X_{ji} + u_i, \tag{6.1}$$

where r is a our reform effect, r_0 is a constant, β is a vector of coefficients, X_j are variables capturing study characteristics (including the standard error) and u_i is a normal IID error term. The problem is that our matrix of regressors X includes a relatively high number of collected independent variables X_j, and we are uncertain about which subset to use. For our dataset, we collected 45 different characteristics of the primary studies and their respective models. Using all of them would be inefficient, hence, we face the problem of identifying empirically relevant independent variables. With 45 variables, we could estimate 2^{45} variable combinations and build 2^{45} different models, therefore, we wish to apply the most efficient way to determine the most relevant set of variables. Should we apply the frequentist meta-regression, we would need to identify a set of key explanatory variables and a set of other controls about which we have weaker belief. Afterward we would have to sequentially remove insignificant variables one by one using t-tests. Such an approach would be with our number of variables not only cumbersome but most importantly statistically invalid. Ignoring the model uncertainty could lead to biased estimation, misleading inference and prediction. BMA provides a way around the model uncertainty - a mechanism that estimates models for all possible combinations of X_j and calculates a weighted average over all of them.

Another reason why Bayesian approach is more appealing is the model interpretation. A standard frequentist procedure returns coefficients and their p-values that indicate the strength of the evidence against the null hypothesis (that particular coefficient is zero). Evidence suggests (Raftery 1995) that p-values provide misleading results especially in large samples, and can lead to inconsistent conclusions as they do not take model uncertainty into effect. On the other hand, BMA returns posterior effect probabilities that i) take uncertainty into account, and ii) enable to distinguish between rejecting the null hypothesis due to insufficient data and due to evidence *for* the null (Hoeting *et al.* 1999).

How does the Bayesian model averaging work then? Our goal is to assess the size of each variable's effect on reform effects with a view to designing future studies, and to be able to correctly predict the impact of a particular reform. Bayesian estimation views unknown parameters of the model as random variables and uses Bayes' theorem and the law of total probability to derive all the results.

We start with data D, N explanatory variables and K models, where $K = 2^N$, because each variable is either in or out of the model. Before any data are observed, our beliefs and uncertainty about the model M_k are represented by a marginal probability $Pr(M_k)$, i.e. the prior probability that M_k is the true model. Having observed the data, the posterior probability for model M_k stems from an extended form of Bayes' theorem, and is given by

$$Pr(M_k|D) = \frac{Pr(D|M_k)Pr(M_k)}{Pr(D)} = \frac{Pr(D|M_k)Pr(M_k)}{\sum_{l=1}^{K} Pr(D|M_l)Pr(M_l)}, \tag{6.2}$$

where

$$Pr(D|M_k) = \int Pr(D|\beta_k, M_k)Pr(\beta_k|M_k)d\beta_k \tag{6.3}$$

is the integrated likelihood of model M_k, β_k is the vector of parameters of model M_k, $Pr(\beta_k|M_k)$ is the prior density of β_k under model M_k, and $Pr(D|\beta_k, M_k)$ is the likelihood in conventional form. Conditional probability $Pr(M_k|D)$ in (6.2) is the likelihood of model M_k occurring given data D, and is also called the Posterior Model Probability (PMP) as it represents the degree of belief in M_k having accounted for D. (Hoeting *et al.* 1999) Taking into account all possible models M_k, the posterior distribution of our effect sizes r given data D is

$$Pr(r|D) = \sum_{k=1}^{K} Pr(r|M_k, D)Pr(M_k|D), \tag{6.4}$$

where $M_1, ..., M_K$ are the models considered. The overall posterior distribution equation (6.4) shows an average of the posterior distributions under each of the models considered, weighted by their posterior model probability.

Averaging across the model space provides better average predictive ability than any other single model M_j and thus we can derive the weighted expected value of r, i.e. the posterior mean, as

$$E[r|D] = \sum_{k=0}^{K} \hat{r}_k Pr(M_k|D), \tag{6.5}$$

where $\hat{r}_k = E[r|D, M_k]$, i.e. the expected reform effect given data D and model M_k. The corresponding posterior variance of the effect size r is then:

$$Var[r|D] = \sum_{k=0}^{K} (Var[r|D, M_k] + \hat{r}_k^2)Pr(M_k|D) - E[r|D]^2. \tag{6.6}$$

To complete the BMA model, we need to define priors on the model space and on the distribution of the coefficients β as well as we need the ability to calculate (or approximate) all the integrals and explore the model space, therefore we address these issues in the following section.

6.2 Implementing Bayesian Model Averaging

In our case, Bayesian estimation is very attractive and brings us a number of advantages concerning the model uncertainty, yet we have to tackle several difficulties during its implementation.

First, the number of models to be weighted might be very high making the summation in (6.4) infeasible. One way of dealing with this problem is to average over a subset of models that receive more support from the data than the rest.[1] Another way uses a Markov Chain Monte Carlo (MCMC) method to directly approximate (6.4). More precisely, Metropolis-Hastings algorithm simulates a Markov chain whose equilibrium distribution is the desired posterior distribution. For a detailed description of the MCMC method see Hoeting *et al.* (1999) and a manual by StataCorp (2017).

Second, the integrals in (6.2) can be hard to compute and require various approximations. Fortunately, for linear regression models, that we use in our analysis, closed form integrals of marginal likelihood $Pr(D|M_k)$ are available.

Third, we need to specify both parts of the prior - prior distribution over all parameters in all competing models and prior probability of each model - which might become challenging. Before we look into the data, we formulate our prior beliefs on coefficients into a normal distribution with a specified mean and variance (Zeugner *et al.* 2015). If we do not have much information about them, a prior mean of zero is a conservative choice. Their variance is then defined according to Zellner's g $(\beta|g(0, \sigma^2(\frac{1}{g}X'X)^{-1}))$, which embodies how sure we are that the coefficients are zero. In our baseline specification, we apply a popular option for g-prior - unit information prior (UIP), which sets $g = N$ for all models. This means that the prior has approximately the same information as has one observation. As a robustness check, we employ a *"BRIC"* prior, which sets $g = max(N, K^2)$, which makes PMPs behave like the Bayesian information criteria or the risk inflation criteria (Zeugner *et al.* 2015).

When we have beliefs about the importance of a particular model for a model structure, prior probability on model M_k can be written as:

$$Pr(M_k) = \prod_{j=1}^{p} \pi_j^{\delta_{kj}}(1-\pi_j)^{1-\delta_{kj}}, \tag{6.7}$$

where $\pi_j \in [0, 1]$ is the prior probability that $\beta_j \neq 0$ in a regression model, and δ_{ij} is an indicator of whether or not variable j is included in the model M_k. (Hoeting *et al.* 1999) When there is little knowledge about the importance of variables and models, the assumption that all models are equally likely is a "neutral" choice. In such case, the prior model probability would be set as uniform, i.e. all are equally likely to be correct. Eicher *et al.* (2011) Incorporating informative prior distributions, however, provides improved predictive performance.

[1] This approach is called the Occam's window method; for details see Madigan & Raftery (1994).

All potential explanatory variables for our regression are listed in Chapter 3. Variable *ref_type* is categorical, therefore it is transformed into four dummy variables - *ref_type1* for labour market reforms, *ref_type2* for fiscal and institutional reforms, *ref_type3* for financial market and innovation reforms, and *ref_type4* for welfare state reforms. In order to avoid the dummy variable trap, we omit *ref_type2* which then poses as a reference value. The same applies to variables describing the type of the reform measure, where we omit *ebrd*. In addition, we exclude dummies for specific parts of a reform index (*lie*, *lii*, *lip*), because these characteristics are captured by the reform type variables. Some of the variables have too little variance and they would be automatically omitted, therefore they are not included in the regression. These include dummies *empl*, *service*, for the long-run estimates fdi and *comb*.

In order to avoid multicollinearity, we compute a correlation matrix (reported graphically in Appendix C) for all study characteristics we intend to use. Figure D.2 depicts the Pearson correlations on a colour scale where yellow denotes a total positive correlation and purple a total negative correlation. The highest correlation is between the total citations from Google Scholar and citations per year, thus we keep in the regression only the latter. Three pairs - $stabil - nstabil$, $inst - ninst$, $fact - nfact$ - are strongly correlated as well, hence, in each case we include only the number of controls. A strong negative correlation is also between variable *start* and *tspan*, however, both of them remain in the regression model.

The final version of our regression model 6.1 includes 35 explanatory variables that capture characteristics of the estimate itself and the study that reported it. Before we run the BMA procedure, we manually weigh the observations by the number of estimates from the same study so as to have a comparable result to the OLS regression from Chapter 5. For the BMA procedure, we choose UIP and $uniform$ priors and run the model using the BMS package in R. Additionally, we perform a robustness check using alternative priors.

The reported results of BMA analysis include the model inclusion figures and four statistic measures: posterior inclusion probability, weighted posterior mean, weighted posterior variance and conditional posterior sign, for all explanatory variables considered.

Posterior Inclusion Probability (PIP) is the posterior probability that a particular variable is included in the model. Mathematically, it is the sum of PMPs for all models which include the particular variable. PIP represents the importance of the variable in explaining the data - the higher it is, the more important is the variable in explaining the heterogeneity (high posterior probabilities of being included are considered as robust).

Weighted posterior mean calculated as in (6.5) is the model averaged parameter estimate. It is derived from the individual model estimates that are weighted by their posterior model probabilities. The models where the variable is not contained are also included in the average (with the parameter estimate being zero).

Weighted posterior variance calculated as in (6.6) incorporates not only the weighted average of the estimated variances of the individual models but also the weighted variance in estimates of the parameters β_j across different models. This means that even if we have highly precise estimates in all the models, we might end up with considerable uncertainty about the parameter if those estimates are very different across specifications.

Conditional posterior sign describes the sign certainty and is defined as the posterior probability of a positive coefficient expected value conditional on inclusion Zeugner *et al.* (2015). In other words, if the value of this statistic is one for some variable, it means that the expected value of the coefficient is positive in all models that include this particular variable. In contrast, if the conditional posterior sign is close to zero, the expected value of the coefficient is negative in most of the encountered models.

6.3 Results

Figures E.1 and E.2 report our results of the BMA analysis in terms of model inclusion of different independent variables. In the figures, different regression models are shown in columns and are scaled by their posterior model probabilities on the horizontal axis - the wider is the column, the more likely the model. The individual independent variables are sorted according to their posterior inclusion probability, with the variables with the highest PIP on top. If cells in the figure are not filled with any colour, the corresponding variable is not included in that particular model. On the other hand, if the cell is coloured, it means that the variable is included in the model. More specifically, the red cells represent variables with a positive coefficient and the blue cells with a negative one in the regression. We can see that approximately a half of the variables appears in the best models for the short run, and about a third for the long run. Within each row, the sign of any variable remains the same, meaning that the signs of the estimated parameters are robust to including other controls.

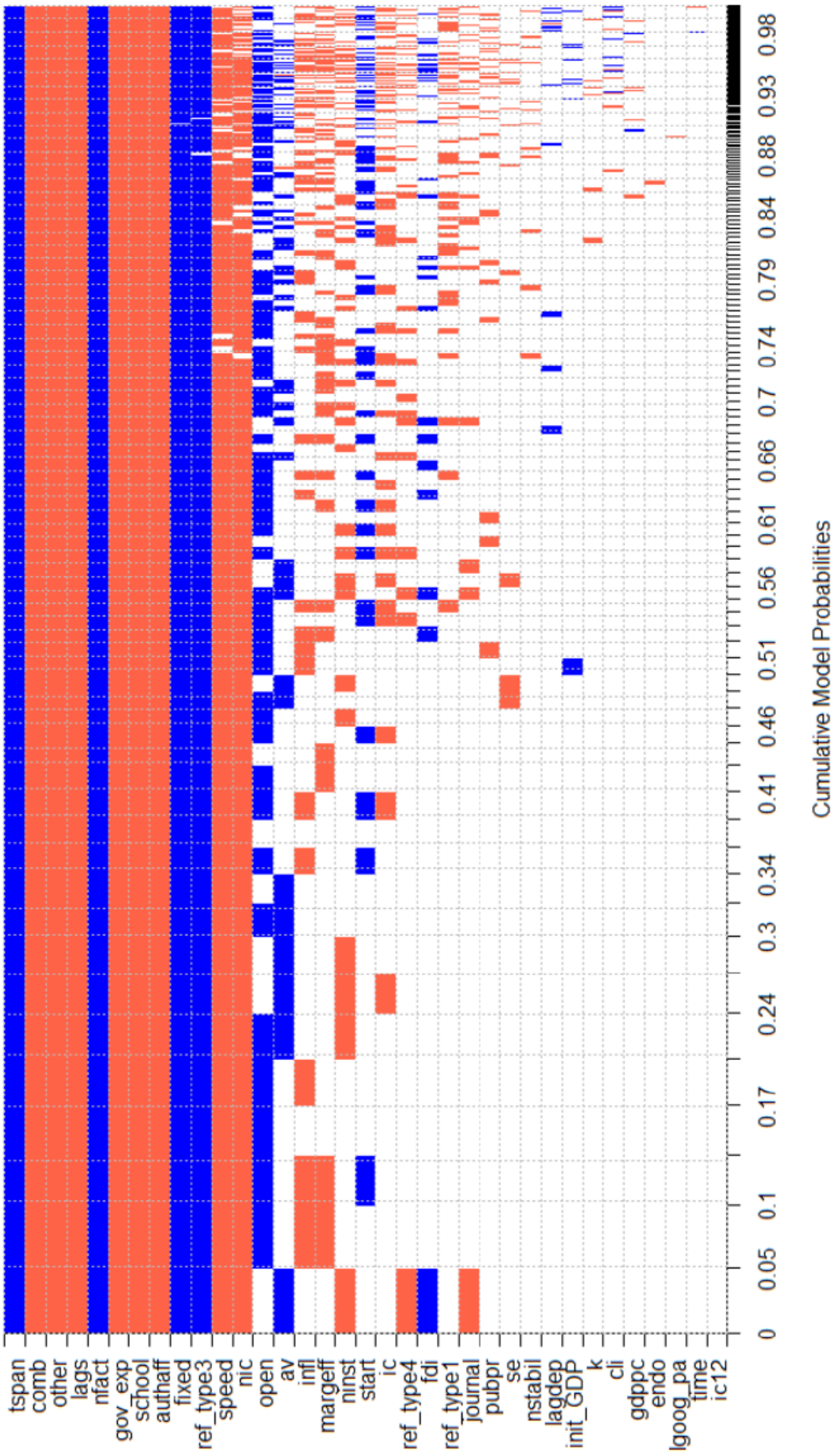
The numerical results of BMA for the short run and long run are reported in tables in Appendix C and further details about the estimation including the diagnostic plots are in Appendix D. The results give information about the PIP, posterior mean, posterior standard deviation and conditional posterior sign for all variables. While the interpretation of the last three is clear and analogical to coefficients and standard errors in standard regression, in case of posterior inclusion probability we follow effect-thresholds described by Kass & Raftery (1995) in order to assess the importance of each variable. Kass & Raftery (1995) considers effect with posterior inclusion probability between 0.5 and 0.75 as weak, between 0.75 and 0.95 as substantial, between 0.95 and 0.99 as strong, and decisive if it exceeds 0.99. We use this scale when interpreting the estimated regression coefficients.

We sort the variables into five categories as in Chapter 3, namely data and methods, type of reform measure, measure of dynamics, specification characteristics and publication characteristics. In each group, there are study characteristics that influence the estimates and help to explain the variation among the reform effects.

The choice of data and methodology systematically affect the reported estimates of both the short run and long run reform effects. Looking first at the short run effects, if the researchers control for the country-specific heterogeneity using fixed effects, the reform effects are 0.12 larger. For the long run effects, researchers using panel data tend to obtain estimates of the reform effect 0.09 smaller than those using cross-sections. Further, the usage of fixed effects and the number of explanatory variables in the regression are the most useful variables when explaining the effect variance.

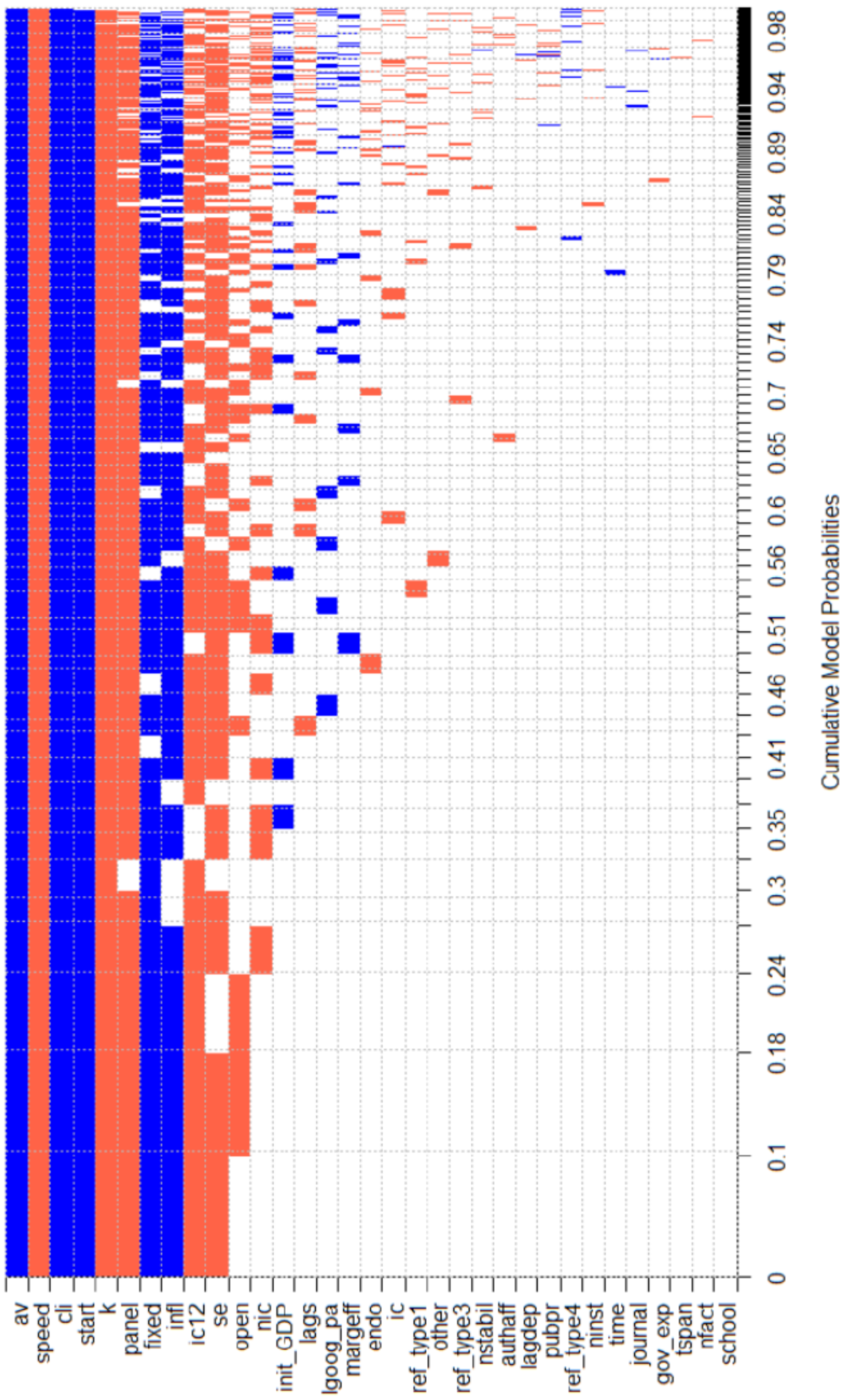
The reform measure, that the researcher chooses, also has a strong influence on the resulting reform effect. The short run effects are affected the most if the chosen measure is other than EBRD index or CLI (the posterior mean is -0.28). In case a combination of EBRD and CLI index is used, the reform effect is again lower. The *ref_type3* variable corresponding to the financial market and innovation reforms has a strong positive effect on the final reform effect compared to the fiscal reforms, contrarily to the labour market and welfare state reforms. A strong to decisive positive effect on the long run estimates has the usage of CLI index as a reform measure and incorporation of an average of components of reform index into the regression. In this case, all three dummies

Figure 6.1: Bayesian Model Averaging, Model Inclusion (Short Run)



Notes: The dependent variable is the reform effect, i.e. the partial correlation coefficient, weighted by the inverse of the number of estimates reported per study. On the horizontal axis, the figure shows the best models, scaled by their PMPs. The explanatory variables are plotted on the vertical axis. Blue colour corresponds to a positive coefficient, red to a negative coefficient, and white to non-inclusion of the particular variable.

Figure 6.2: Bayesian Model Averaging, Model Inclusion (Long Run)



Notes: The dependent variable is the reform effect, i.e. the partial correlation coefficient, weighted by the inverse of the number of estimates reported per study. On the horizontal axis, the figure shows the best models, scaled by their PMPs. The explanatory variables are plotted on the vertical axis. Blue colour corresponds to a positive coefficient, red to a negative coefficient, and white to non-inclusion of the particular variable.

for the type of reform have posterior inclusion probability of less that 0.1.

Regarding the general design of the reform equation, it matters for the estimated effect whether the researchers control for the dynamics. We know that if they want to estimate the short run effects, they need to control for the dynamics. With respect to this fact, it comes as a no surprise that if studies control for the reform speed or include lagged reform variables, they get lower reform effects. In case of the long run effects, the only variable with decisive impact is *speed*, which inclusion in the model makes the estimates 0.28 smaller.

When the authors of primary studies control for schooling or government expenditures, they report short-run estimates lower by more than 0.20. Other decisive factor, that affects the short-run estimates is the number of controls for the factors of production, though in the positive direction (posterior mean of 0.10). It seems that other controls do not affect the resulting reform effects; only the number of controls for initial conditions has a substantial effect and openness has a weak one. When estimating the long run reform effect, it is important to control for the stabilization as well - if authors do not control for inflation, they are likely to report smaller estimates. The only other study characteristics that has at least a weak impact on the long-run reform effects is controlling for the initial conditions using principal components. Studies that do so report long-run effects by 0.08 lower than studies that do not control for initial conditions at all.

Regarding the publication and other study characteristics, the fact that the paper was published in a refereed journal is not important for the reported reform effect. Neither is important the number of citations, and the author affiliation for the long run effects. The short run effects, however, tend to be 0.11 smaller when estimated by the authors entirely from academia.

The choice of priors substantially affects the results of BMA, therefore we report an additional set of results in Appendix E, in which we use alternative priors - *"random"* model prior and *"BRIC"* Zellner's g. The results obtained are roughly similar to the ones from our baseline specification in terms of both the importance of the study characteristics and the direction of their effects, with only small differences in the posterior inclusion probabilities and effect magnitudes. We can therefore conclude that our results are robust to changes in priors.

6.4 Best-Practice Estimates

Now we know that the most heterogeneity is caused by the choice of data and reform measure, but we would also like to determine the size of the effects corrected for publication bias and study heterogeneity. We compute a synthetic mean estimate of the reform effect from a hypothetical study that uses appropriate methodology and data. Hence, it involves defining best practice in the estimation of reform effect, which is the most subjective part of the analysis, because different researchers may have different opinions of what the best practice is. Nevertheless, for each explanatory variable we select a preferred value or a sample mean if we have no preference and compute the reform effect as a linear combination of all the regression parameters.

In our best practice study, we prefer to use panel data, because they allow to address potential endogeneity and time dynamics. Both of these statistical issues should be controlled for in our synthetic study, along with the country heterogeneity (value of the fixed effects dummy is set to one). As for the reform measure, we prefer using other measures of the reform than EBRD and CLI, because these are reported for transition economies only. We would recommend controlling for the initial conditions, stabilization, inflation, factors of production and institutional development, therefore we plug the sample maxima for all variables that correspond to their number of types of controls. Finally, we plug in sample maximum for the number of citations per year and as we prefer studies published in academic journals, we plug in 1 for the corresponding dummy. All other variables are set to their sample means.

In line with the defined best practice, we calculate synthetic estimates differentiating the short run and the long run as well as different types of reforms. Results with the improved estimates of the reform effect are reported in Table 6.1. Corresponding 95% confidence intervals are constructed using study-level clustered standard errors estimated by OLS. In general, all of the best-practice estimates are negative yet insignificant, except for the welfare state reform effect that is significant and negative (-0.39). Problem in this case is the lack of observations on the short-run effect of welfare state reforms, thus we need to treat this result with caution. Compared with the simple means reported in Table 4.1, all of the best-practice estimates are smaller, which captures both publication bias and study heterogenity. As for the long-run reform effects, all estimated effects are positive, however, only the effect of fiscal and institutional, and innovation and financial market reforms are statistically significant at 5% level. The magnitude of the best-practice effects is slightly larger than the one of the simple means. Finally, the long-run effect of reforms regardless of their scope is positive and significant as well.

In order to assess the magnitude of the reform effects, we follow guidelines by Doucouliagos (2011) described in Chapter 4. All the long-run estimates fall into the category of small effects, only the effect of fiscal reforms is marginally above the threshold set for medium effect. Concerning the short-run effects, the overall effect along with the effect of innovation and financial market reforms is small, the effect of labour market and fiscal reforms is of medium size, and the effect of welfare state reforms is strong.

Our best-practice reform-growth effects are measured in partial correlation coefficients, because other measures, such as elasticities, are not available. The problem is that it is difficult to imagine and interpret, what the values of the partial correlations mean for the growth. With this purpose, we calculate the short- and long-run elasticities of growth for each type of reform reported in Table 6.2. For this calculation, we follow the same procedure as described in Babecky & Havranek (2014) - we use dataset from Havranek *et al.* (2015) and regress the elasticities reported there on partial correlation coefficients. This tells us more about the relationship between the two variables. Then we use the estimated coefficients and using our best-practice estimates, we calculate the elasticities of growth with respect to reforms. Our results show that the short-run elasticity of growth is -0.51 and the long-run elasticity 0.49.

Table 6.1: Estimating the Average Reform Effect

	Short run			Long run		
Reform type	Est. effect	95% CI		Est. effect	95% CI	
All types of reforms						
	-0.167	-0.433	0.099	0.154	0.069	0.240
Labour market reforms						
	-0.274	-0.589	0.041	0.090	-0.063	0.244
Fiscal reforms and institutions						
	-0.185	-0.451	0.081	0.173	0.081	0.265
Innovation and financial markets						
	-0.104	-0.373	0.164	0.122	0.003	0.241
Welfare state reforms						
	-0.389	-0.714	-0.065	0.121	-0.065	0.306

Notes: Est. effect = estimated reform effect represented by the partial correlation coefficient; CI = confidence interval.

Table 6.2: Estimating the Elasticity of Growth

Reform type	Short run	Long run
All types of reforms	-0.507	0.494
Labour market reforms	-0.614	0.430
Fiscal reforms and institutions	-0.525	0.513
Innovation and financial markets	-0.444	0.462
Welfare state reforms	-0.729	0.461

Chapter 7

Conclusion

In this thesis, we conduct a systematic review of the literature that estimates the effect of structural reforms on economic performance in European countries. We apply modern meta-analytical methods to (i) estimate the most precise reform effects, (ii) explain why the effects vary and (iii) examine the differences in reactions of the economic growth to different reforms. We collect 889 reform effects from 90 studies and employ partial correlation coefficients in order to standardize them. Controlling for publication bias and differences in study quality, we confirm the long-run positive effect of reforms on economic performance, yet we do not find enough evidence for the hypothesis that reforms bring costs in the short run. Our results further indicate that fiscal and institutional reforms are more beneficial than other types of reforms as opposed to the labour market reforms that appear the most costly.

We test the reform effects for publication selection in order to avoid drawing false conclusions from potentially biased literature. For this purpose, we use informal yet instructive funnel plots and formal regression-based tests, which we estimate using four estimation techniques (namely, OLS, WLS, fixed effects, and instrumental variables estimation). All regressions are corrected for heterodskedasticity using study-level clustered standard errors. We find little evidence for publication selection - positive for the overall long-run effects and negative for the short-run labour market reform effects, which means that negative long-run reform effects and positive short-run effects of labour market reforms are systematically underreported.

To uncover the link between the study design and the reported estimates, we use Bayesian model averaging, which provides a solution to uncertainty in model selection. Our results show that data choice and methodology matter for the reported reform-growth effects. Researchers that choose panel data over cross-sections tend to obtain smaller long-run estimates. Next, accounting for the country-heterogeneity in the reform-growth regressions is important and yields larger reform estimates. The type of reform measure is also a relevant factor that determines how strong the estimated reform effect is. If researchers use other measures of reform than EBRD index or cumulative liberalization index, that are standard for assessing the reforming effort in transition economies, their estimated impact of the reform is smaller. We check for the robustness of our results by changing priors in the BMA procedure and we find that the results of this robustness check are in line with our baseline model.

Using the estimated regression coefficients from BMA, we construct synthetic estimates of the typical reaction of economic growth to reforms. In order to do so, we define a best-practice model with the most preferred methodology. This approach yields statistically insignificant estimates of the short run reform effect ranging from -0.10 to -0.39, which could be described as a small to medium effect. The long-run reform effects are mostly significant and with values between 0.09 and 0.17, which is classified as a small effect. Among the different types of reforms, fiscal and institutional reforms show more positive effects on growth than any other type, however, the reforms of financial markets and innovation spurring reforms seem to have the lowest short-run costs.

Meta-analysis is a powerful tool that allows to combine the knowledge of a substantial part of the research field. At the same time, it is of crucial importance that the policymakers understand how the reforms influence economic performance, which type of reforms brings the lowest costs and which the largest gains. We hope that this meta-analysis might, therefore, (i) shed some light on the significance of the effect of structural reforms on growth, (ii) provide more structured evidence and more precise and accurate estimates (iii) improve understanding of the impact different reforms have on the economy.

Bibliography

Abed, G. T. & H. R. Davoodi (2002): "Corruption, Structural Reforms, and Economic Performance in the Transition Economies." In S. Gupta & G. T. Abed (editors), "Governance, Corruption, and Economic Performance," chapter 23, pp. 489–537. Washington, D.C.: International Monetary Fund.

Ahrens, J. & M. Meurers (2002): "How governance affects the quality of policy reform and economic performance: new evidence for economies in transition." *Journal for Institutional Innovation, Development and Transition* **6**: pp. 35–56.

Aloe, A. M. & C. G. Thompson (2013): "The synthesis of partial effect sizes." *Journal of the Society for Social Work and Research* **4(4)**: pp. 390–405.

Apolte, T. (2011): "Democracy and prosperity in two decades of transition." *Economics of Transition* **19(4)**: pp. 693–722.

Arnold, J. (2008): "Do tax structures affect aggregate economic growth?: Empirical evidence from a panel of OECD countries." *OECD Economics Department Working Papers* **(643)**.

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

Åslund, A., P. Boone, S. Johnson, S. Fischer, & B. W. Ickes (1996): "How to Stabilize: Lessons from Post-communist Countries." *Brookings Papers on Economic Activity* **27(1)**: pp. 217–313.

Aziz, M. J. & M. R. F. Westcott (1997): "Policy Complementarities and the Washington Consensus." *IMF Working Paper No. 97/118* .

Babecky, J. & N. F. Campos (2011): "Does reform work? An econometric survey of the reform–growth puzzle." *Journal of Comparative Economics* **39(2)**: pp. 140–158.

Babecky, J. & T. Havranek (2014): "Structural Reforms and Growth in Transition." *Economics of Transition* **22(1)**: pp. 13–42.

Badinger, H. (2005): "Growth effects of economic integration: Evidence from the EU member states." *Review of World Economics* **141(1)**: pp. 50–78.

Barone, G. & F. Cingano (2011): "Service regulation and growth: Evidence from OECD countries." *The Economic Journal* **121(555)**: pp. 931–957.

Bassanini, A., S. Scarpetta, & P. Hemmings (2001): "Economic Growth: The Role of Policies and Institutions: Panel data. Evidence from OECD countries." *OECD Economics Department Working Papers* **(283)**.

Beck, T. & L. Laeven (2006): "Institution building and growth in transition economies." *Journal of Economic Growth* **2(11)**: pp. 157–186.

Berg, A., E. Borensztein, R. Sahay, & J. Zettelmeyer (1999): "The Evolution of Output in Transition Economies: Explaining the Differences." *IMF Working Paper No. 99/73* .

Berti, K., E. Meyermans *et al.* (2017): "Maximising the impact of labour and product market reforms in the euro area – sequencing and packaging." *Quarterly Report on the Euro Area (QREA)* **16(2)**: pp. 7–19.

Bojnec, Š. & I. Lejko (2012): "Internationalization and economic growth in the new member states of the European Union." *Ekonomický časopis* **60(4)**: pp. 335–348.

Borenstein, M., L. V. Hedges, J. P. Higgins, & H. R. Rothstein (2010): "A basic introduction to fixed-effect and random-effects models for meta-analysis." *Research Synthesis Methods* **1(2)**: pp. 97–111.

Bosma, N., J. Content, M. Sanders, & E. Stam (2017): "Time series and panel data analysis of GEI and growth performance indicators." *Unpublished* .

Bouis, R., O. Causa, L. Demmou, R. Duval, & A. Zdzienicka (2012): "The Short-Term Effects of Structural Reforms." *OECD Economic Department Working Papers* **(949)**.

Bouis, R. & R. Duval (2011): "Raising Potential Growth After the Crisis: A Quantitative Assessment of the Potential Gains from Various Structural Reforms in the OECD Area and Beyond." *OECD Economifc Department Working Papers* **(835)**.

Böwer, U. & A. Turrini (2010): "EU accession: A road to fast-track convergence?" *Comparative Economic Studies* **52(2)**: pp. 181–205.

Bradley, M. T. & R. D. Gupta (1997): "Estimating the Effect of the File Drawer Problem in Meta-Analysis." *Perceptual and Motor Skills* **85(2)**: pp. 719–722.

Brissimis, S. N., M. D. Delis, & N. I. Papanikolaou (2008): "Exploring the nexus between banking sector reform and performance: Evidence from newly acceded EU countries." *Journal of Banking & Finance* **32(12)**: pp. 2674–2683.

Burns, P. B., R. J. Rohrich, & K. C. Chung (2011): "The levels of evidence and their role in evidence-based medicine." *Plastic and Reconstructive Surgery* **128(1)**: p. 305.

Cacciatore, M., R. Duval, & G. Fiori (2012): "Short-term gain or pain? A DSGE model-based analysis of the short-term effects of structural reforms in labour and product markets." *OECD Publishing* .

Caporale, G. M., C. Rault, A. D. Sova, & R. Sova (2015): "Financial development and economic growth: Evidence from 10 new European Union members." *International Journal of Finance & Economics* **20(1)**: pp. 48–60.

Cappelen, A., F. Castellacci, J. Fagerberg, & B. Verspagen (2003): "The impact of EU regional support on growth and convergence in the European Union." *Journal of Common Market Studies* **41(4)**: pp. 621–644.

Cappellari, L., C. Dell'Aringa, & M. Leonardi (2012): "Temporary employment, job flows and productivity: A tale of two reforms." *The Economic Journal* **122(562)**: pp. 188–215.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *The American Economic Review* **85(2)**: pp. 238–243.

Cerović, B. & A. Nojković (2009): "Transition and Growth: What was Taught and What Happened." *Economic Annals* **54(183)**: pp. 7–31.

Chattoe-Brown, E. (1998): "Just How (Un) realistic Are Evolutionary Algorithms As Representations of Social Processes?" *Journal of Artificial Societies and Social Simulation* **1(3)**.

Christoffersen, P. & P. Doyle (2000): "From Inflation to Growth." *Economics of Transition* **8(2)**: pp. 421–451.

Cieślik, A. & M. Tarsalewska (2013): "Privatization, Convergence, and Growth: For Transition Economies." *Eastern European Economics* **51(1)**: pp. 5–20.

Cingano, F., M. Leonardi, J. Messina, & G. Pica (2016): "Employment protection legislation, capital investment and access to credit: Evidence from Italy." *The Economic Journal* **126(595)**: pp. 1798–1822.

Comunale, M. & F. P. Mongelli (2018): "Who did it? A European Detective Story." *Unpublished* .

Cuadrado-Ballesteros, B. & N. Peña-Miguel (2017): "The Socioeconomic Consequences of Privatization: An Empirical Analysis for Europe." *Social Indicators Research* **139(4)**: pp. 1–21.

Cungu, A. & J. Swinnen (2003): "The impact of aid on economic growth in transition economies: An empirical study." *KU Leuven Discussion Papers 128/2003* .

De Macedo, J. B. & J. O. Martins (2008): "Growth, reform indicators and policy complementarities 1." *Economics of Transition* **16(2)**: pp. 141–164.

De Melo, M., C. Denizer, & A. Gelb (1996): "Patterns of transition from plan to market." *The World Bank Economic Review* **10(3)**: pp. 397–424.

De Melo, M., C. Denizer, A. Gelb, & S. Tenev (2001): "Circumstance and Choice: The Role of Initial Conditions and Policies in Transition Economies." *The World Bank Economic Review* pp. 1–31.

De Melo, M., C. Denizer, & A. H. Gelb (1997): *From Plan to Market–Pattern of Transition.* World Bank Publications.

Denizer, C. (1999): "Stabilization, adjustment, and growth prospects in transition economies." *Policy Research Working Papers, The World Bank* .

Domaç, I., K. Peters, & Y. Yuzefovich (2001): "Does the exchange rate regime affect macroeconomic performance? Evidence from transition economies." *Policy Research Working Papers, The World Bank* .

Doucouliagos, H. (2011): "How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics." *Technical report*, Deakin University.

EBRD (2010): "EBRD Transition Report 2010: Recovery and Reform."

EBRD (2018a): "EBRD Transition indicators methodology."

EBRD (2018b): "EBRD Transition Report 2018-19: Work in Transition."

Eggertsson, G., A. Ferrero, & A. Raffo (2014): "Can structural reforms help Europe?" *Journal of Monetary Economics* **61**: pp. 2–22.

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.

Eicher, T. S. & T. Schreiber (2010): "Structural policies and growth: Time series evidence from a natural experiment." *Journal of Development Economics* **91(1)**: pp. 169–179.

Eschenbach, F. & B. Hoekman (2005): *Services policy reform and economic growth in transition economies, 1990-2004*. The World Bank.

European Commission (2018a): "European Semester: Assessment of progress on structural reforms, prevention and correction of macroeconomic imbalances, and results of in-depth reviews under Regulation (EU) No 1176/2011."

European Commission (2018b): "Structural reforms for economic growth."

Faini, R., J. Haskel, G. B. Navaretti, C. Scarpa, & C. Wey (2004): "Contrasting europe's decline: do product market reforms help?" In "a conference organised by Fondazione Rodolfo Debenedetti, Lecce," volume 19.

Falcetti, E., T. Lysenko, & P. Sanfey (2006): "Reforms and growth in transition: Re-examining the evidence." *Journal of Comparative Economics* **34(3)**: pp. 421–445.

Falcetti, E., M. Raiser, & P. Sanfey (2002): "Defying the odds: Initial conditions, reforms, and growth in the first decade of transition." *Journal of comparative economics* **30(2)**: pp. 229–250.

Fatas, A. (2016): "The agenda for structural reform in europe." *After the Crisis: Reform, Recovery, and Growth in Europe* .

Fetahi-Vehapi, M., L. Sadiku, & M. Petkovski (2015): "Empirical analysis of the effects of trade openness on economic growth: An evidence for south east european countries." *Procedia Economics and Finance* **19**: pp. 17–26.

Fidrmuc, J. (2003): "Economic reform, democracy and growth during post-communist transition." *European journal of political economy* **19(3)**: pp. 583–604.

Fidrmuc, J. & A. Tichit (2004): "Mind the break! accounting for changing patterns of growth during transition." *Technical report*, William Davidson Institute at the University of Michigan.

Fidrmuc, J. & A. Tichit (2009): "Mind the break! accounting for changing patterns of growth during transition." *Economic Systems* **33(2)**: pp. 138–154.

Fischer, K., A. Stiglbauer *et al.* (2018): "Structural reforms for higher productivity and growth." *Monetary Policy and the Economy Q* **2**: pp. 132–152.

Fischer, S. & R. Sahay (2000): "The transition economies after ten years." *Technical report*, National bureau of economic research.

Fischer, S. & R. Sahay (2004): "Transition economies: The role of institutions and initial conditions." In "Calvo Conference, IMF," Citeseer.

Fischer, S., R. Sahay, & C. A. Végh (1996a): "Economies in transition: The beginnings of growth." *The American Economic Review* **86(2)**: pp. 229–233.

Fischer, S., R. Sahay, & C. A. Vegh (1996b): "Stabilization and growth in transition economies: the early experience." *Journal of economic perspectives* **10(2)**: pp. 45–66.

Fischer, S., R. Sahay, & C. A. Vegh (1997): "From transition to market: evidence and growth prospects." In "Lessons from the Economic Transition," pp. 79–101. Springer.

Gavriluţă, A. F., F. Oprea *et al.* (2017): "Fiscal Decentralization Determinants And Local Economic Development In EU Countries." *European Union at crossroads building resilience in times of change* p. 180.

Gelb, A., M. Melo, C. Denizer, & S. Tenev (1999): *Circumstance and choice: The role of initial conditions and policies in transition economies.* The World Bank.

Gillman, M. & M. Kejak (2005): "Contrasting models of the effect of inflation on growth." *Journal of Economic Surveys* **19(1)**: pp. 113–136.

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

Godoy, S. & J. E. Stiglitz (2007): "Growth, initial conditions, law and speed of privatization in transition countries: 11 years later." In "Transition and beyond," pp. 89–117. Springer.

Gomes, S., P. Jacquinot, M. Mohr, & M. Pisani (2013): "Structural reforms and macroeconomic performance in the euro area countries: A model-based assessment." *International Finance* **16(1)**: pp. 23–44.

Griffith, R., R. Harrison, & H. Simpson (2010): "Product market reform and innovation in the EU." *Scandinavian Journal of Economics* **112(2)**: pp. 389–415.

Griffith, R., R. Harrison, H. Simpson *et al.* (2006): "The link between product market reform, innovation and EU macroeconomic performance." *Technical report*, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Harrison, F. (2011): "Getting started with meta-analysis." *Methods in Ecology and Evolution* **2(1)**.

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.

Havrylyshyn, O. & R. Van Rooden (1998): "Recovery and growth in transition economies 1990-97: A stylized regression analysis." *Unpublished* .

Havrylyshyn, O. & R. Van Rooden (2003): "Institutions matter in transition, but so do policies." *Comparative Economic Studies* **45(1)**: pp. 2–24.

Heipertz, M. & M. Ward-Warmedinger (2008): "Economic and social models in Europe and the importance of reform." *Financial Theory and Practice* **32(3)**: pp. 255–287.

Hernández-Catá, E. (1997): "Liberalization and the behavior of output during the transition from plan to market." *Staff Papers* **44(4)**: pp. 405–429.

Heybey, B. & P. Murrell (1999): "The relationship between economic growth and the speed of liberalization during transition." *The Journal of Policy Reform* **3(2)**: pp. 121–137.

Hobza, A., G. Mourre *et al.* (2010): "Quantifying the potential macroeconomic effects of the Europe 2020 strategy: stylised scenarios." *Technical report*, Directorate General Economic and Financial Affairs (DG ECFIN).

Hoeting, J. A., D. Madigan, A. E. Raftery, & C. T. Volinsky (1999): "Bayesian model averaging: a tutorial." *Statistical science* pp. 382–401.

Hopewell, S., K. Loudon, M. J. Clarke, A. D. Oxman, & K. Dickersin (2009): "Publication bias in clinical trials due to statistical significance or direction of trial results." *The Cochrane Library* .

Iradian, G. (2009): "What explains the rapid growth in transition economies?" *IMF Staff Papers* **56(4)**: pp. 811–851.

Jaimovich, N. & S. Rebelo (2017): "Nonlinear effects of taxation on growth." *Journal of Political Economy* **125(1)**: pp. 265–291.

Jona-Lasinio, C. & G. Vallanti (2013): "Reforms, labour market functioning and productivity dynamics: a sectoral analysis for Italy." *Unpublished* .

Josifidis, K., R. Dragutinović-Mitrović, & O. Ivančev (2012): "Heterogeneity of growth in the West Balkans and Emerging Europe: a dynamic panel data model approach." *Panoeconomicus* **59(2)**: pp. 157–183.

Kang, Y.-D., C.-W. Lee, T. H. Oh, H. J. Lee, & J. Kim (2014): "Harmonizing Social Welfare and Economic Growth: Case Studies of European Countries and Implications for Korea." *Unpublished* .

Kass, R. E. & A. E. Raftery (1995): "Bayes factors." *Journal of the American Statistical Association* **90(430)**: pp. 773–795.

Kim, B.-Y. & J. Pirttila (2003): "The political economy of reforms: Empirical evidence from post-communist transition in the 1990s." *BOFIT Discussion Papers 4/2003, Bank of Finland, Institute for Economies in Transition* .

Krueger, G. & M. Ciolko (1998): "A note on initial conditions and liberalization during transition." *Journal of Comparative Economics* **26(4)**: pp. 718–734.

Lawson, C. & H. Wang (2005): "Economic transition in central and Eastern Europe and the Former Soviet Union: which policies worked?" *Technical report*, University of Bath, Department of Economics and International Development.

Loungani, P. & N. Sheets (1997): "Central bank independence, inflation, and growth in transition economies." *Journal of Money, Credit, and Banking* pp. 381–399.

Madigan, D. & A. E. Raftery (1994): "Model selection and accounting for model uncertainty in graphical models using Occam's window." *Journal of the American Statistical Association* **89(428)**: pp. 1535–1546.

Magazzino, C. (2014): "Government size and economic growth in Italy: an empirical analyses based on new data (1861-2008)." *Unpublished* .

Masuch, K., E. Moshammer, & B. Pierluigi (2016): "Institutions and growth in Europe." *Unpublished* .

Merlevede, B. (2003): "Reform reversals and output growth in transition economies." *Economics of Transition* **11(4)**: pp. 649–669.

Mickiewicz, T. (2005a): "Is the link between reforms and growth spurious? A comment." *Unpublished* .

Mickiewicz, T. (2005b): "Post-communist recessions re-examined." In "Economic Transition in Central Europe and the Commonwealth of Independent States," pp. 99–118. Springer.

Nath, H. K. (2009): "Trade, foreign direct investment, and growth: Evidence from transition economies." *Comparative Economic Studies* **51(1)**: pp. 20–50.

Neyapti, B. & N. Dincer (2005): "Measuring the quality of bank regulation and supervision with an application to transition economies." *Economic Inquiry* **43(1)**: pp. 79–99.

Nicoletti, G. & S. Scarpetta (2003): "Regulation, productivity and growth: Oecd evidence." *Economic policy* **18(36)**: pp. 9–72.

OECD (2018): "OECD Economic Surveys: European Union." **2018**.

Pääkkönen, J. (2010): "Economic freedom as driver of growth in transition." *Economic Systems* **34(4)**: pp. 469–479.

Pelipas, I. & A. Chubrik (2008): "Market reforms and growth in post-socialist economies: Evidence from panel cointegration and equilibrium correction model." *Unpublished* .

Piculescu, V. (2003): "Direct and feedback effects on economic and institutional developments in transition: a path analysis approach." *Technical report*, Goteborg University, Mimeo.

Polanec, S. (2004): "Convergence at Last?: Evidence from Transition Countries." *Eastern European Economics* **42(4)**: pp. 55–80.

Radulescu, R. & D. Barlow (2002): "The relationship between policies and growth in transition countries." *Economics of Transition* **10(3)**: pp. 719–745.

Radziwill, A. & P. Smietanka (2009): "Eu's eastern neighbours: Institutional harmonisation and potential growth bonus." *Unpublished* .

Raftery, A. E. (1995): "Bayesian Model Selection in Social Research." *Sociological Methodology* **25**: pp. 111–163.

Raimbaev, A. (2011): "The case of transition economies: what institutions matter for growth?" *Technical report*, EERI Research Paper Series.

Rapacki, R., M. Próchniak *et al.* (2009): "The EU enlargement and economic growth in the CEE new member countries." *Technical report*, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Rodrik, D. (2005): "Why we learn nothing from regressing economic growth on policies." *Harvard University* .

Romero-Avila, D. & R. Strauch (2008): "Public finances and long-term growth in Europe: Evidence from a panel data analysis." *European Journal of Political Economy* **24(1)**: pp. 172–191.

Rosenthal, R. (1979): "The file drawer problem and tolerance for null results." *Psychological bulletin* **86(3)**: p. 638.

Sachs, J. D. (1996): "The transition at mid decade." *The American Economic Review* **86(2)**: pp. 128–133.

Selowsky, M. & R. Martin (1997): "Policy performance and output growth in the transition economies." *The American Economic Review* **87(2)**: pp. 349–353.

Shin, I.-S. (2017): "Recent Research Trends in Meta-analysis." *Asian Nursing Research* **11(2)**: pp. 79–83.

Staehr, K. (2005): "Reforms and Economic Growth in Transition Economies: Complementarity, Sequencing and Speed." *The European Journal of Comparative Economics* **2(2)**: pp. 177–202.

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.

StataCorp (2017): *Stata Bayesian Analysis Reference Manual: Release 15.* Stata Press.

Stuckler, D., L. King, & G. Patton (2009): "The social construction of successful market reforms." .

Sutton, A. J., S. Duval, R. Tweedie, K. R. Abrams, & D. R. Jones (2000): "Empirical assessment of effect of publication bias on meta-analyses." *Bmj* **320(7249)**: pp. 1574–1577.

Tausch, A. (2005): "World Bank Pension Reforms and Development Patterns in the World System and in the Wider Europe: A 109 Country Investigation Based on 33 Indicators of Economic Growth, and Human, Social and Ecological Well-Being, and a European Regional Case Study." *Unpublished* .

Thornton, A. & P. Lee (2000): "Publication bias in meta-analysis: its causes and consequences." *Journal of clinical epidemiology* **53(2)**: pp. 207–216.

Wolf, H. C. *et al.* (1999): *Transition strategies: choices and outcomes.* International Finance Section.

Zeugner, S., M. Feldkircher *et al.* (2015): "Bayesian model averaging employing fixed and flexible priors: The BMS package for R." *Journal of Statistical Software* **68(4)**: pp. 1–37.

Appendix A

List of Primary Studies

Table A.1: List of Primary Studies

Åslund *et al.* (1996)	Fischer & Sahay (2000)
Abed & Davoodi (2002)	Fischer & Sahay (2004)
Ahrens & Meurers (2002)	Gavriluţă *et al.* (2017)
Apolte (2011)	Gelb *et al.* (1999)
Arnold (2008)	Gillman & Kejak (2005)
Aziz & Westcott (1997)	Godoy & Stiglitz (2007)
Badinger (2005)	Griffith *et al.* (2006)
Barone & Cingano (2011)	Griffith *et al.* (2010)
Bassanini *et al.* (2001)	Havrylyshyn & Van Rooden (1998)
Beck & Laeven (2006)	Havrylyshyn & Van Rooden (2003)
Berg *et al.* (1999)	Hernández-Catá (1997)
Berti *et al.* (2017)	Heybey & Murrell (1999)
Brissimis *et al.* (2008)	Christoffersen & Doyle (2000)
Bojnec & Lejko (2012)	Iradian (2009)
Bosma *et al.* (2017)	Jaimovich & Rebelo (2017)
Bouis *et al.* (2012)	Jona-Lasinio & Vallanti (2013)
Böwer & Turrini (2010)	Josifidis *et al.* (2012)
Caporale *et al.* (2015)	Kang *et al.* (2014)
Cappelen *et al.* (2003)	Kim & Pirttila (2003)
Cappellari *et al.* (2012)	Krueger & Ciolko (1998)
Cerović & Nojković (2009)	Lawson & Wang (2005)
Cieślik & Tarsalewska (2013)	Loungani & Sheets (1997)
Cingano *et al.* (2016)	Magazzino (2014)
Comunale & Mongelli (2018)	Masuch *et al.* (2016)
Cuadrado-Ballesteros & Peña-Miguel (2017)	Merlevede (2003)
Cungu & Swinnen (2003)	Mickiewicz (2005b)
De Macedo & Martins (2008)	Mickiewicz (2005a)

Continued on the next page

Continued: List of Primary Studies

Continued from previous page

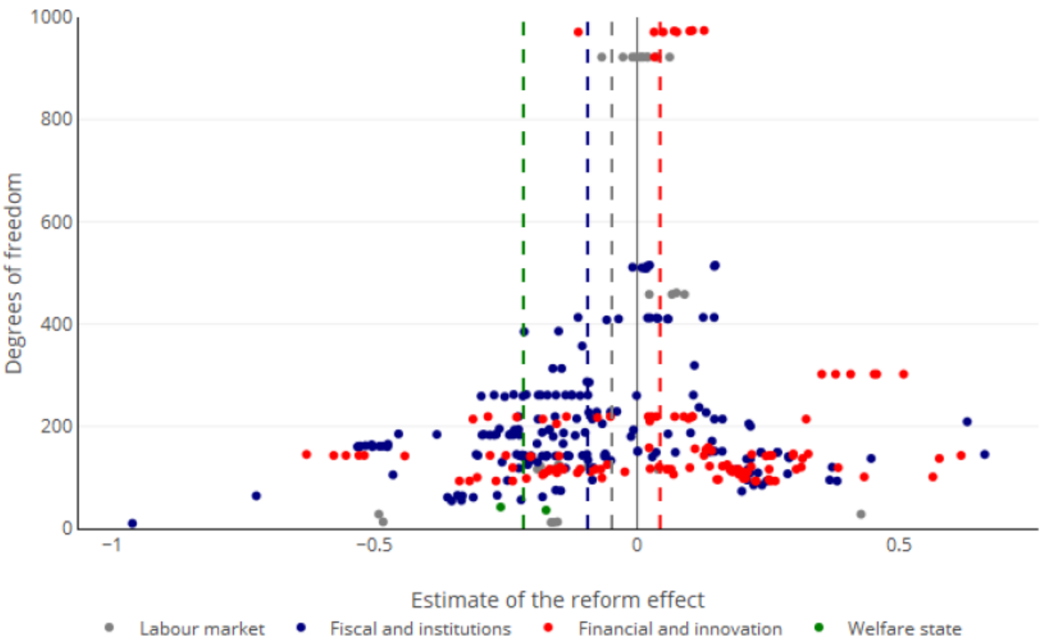
De Melo *et al.* (1996)	Nath (2009)
De Melo *et al.* (1997)	Neyapti & Dincer (2005)
De Melo *et al.* (2001)	Nicoletti & Scarpetta (2003)
Denizer (1999)	Pääkkönen (2010)
Domaç *et al.* (2001)	Pelipas & Chubrik (2008)
Eicher & Schreiber (2010)	Piculescu (2003)
Eschenbach & Hoekman (2005)	Polanec (2004)
Faini *et al.* (2004)	Radulescu & Barlow (2002)
Falcetti *et al.* (2002)	Radziwill & Smietanka (2009)
Falcetti *et al.* (2006)	Raimbaev (2011)
Fatas (2016)	Rapacki *et al.* (2009)
Fetahi-Vehapi *et al.* (2015)	Romero-Avila & Strauch (2008)
Fidrmuc (2003)	Sachs (1996)
Fidrmuc & Tichit (2004)	Selowsky & Martin (1997)
Fidrmuc & Tichit (2009)	Staehr (2005)
Fischer *et al.* (1996a)	Stuckler *et al.* (2009)
Fischer *et al.* (1996b)	Tausch (2005)
Fischer *et al.* (1997)	Wolf *et al.* (1999)

Notes: The search for primary studies was terminated in June 2018.

Appendix B

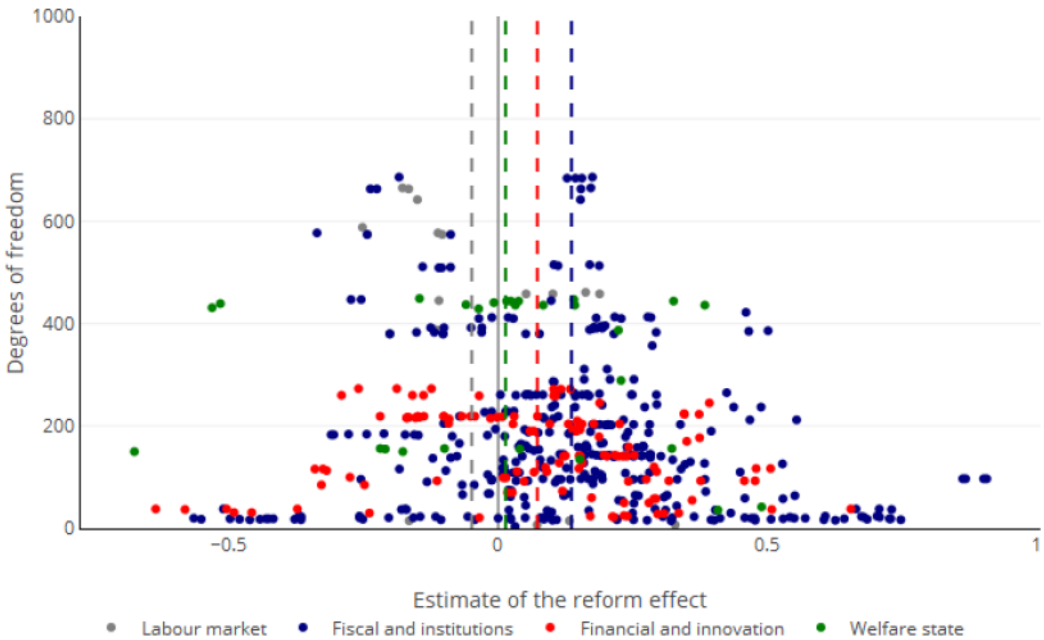
Collected Reform Effects by the Reform Type

Figure B.1: Collected Reform Effects for the Short Run by the Reform Type



Notes: The vertical axis shows the degrees of freedom used for estimation of the reform effect in the primary study; the horizontal axis shows the short run reform effect represented by the partial correlation coefficient. The solid line denotes zero; the dashed lines denote the arithmetic means of the short run estimates by the reform type. The plot shows that the majority of estimates predicts a negative short-run impact of the reforms.

Figure B.2: Collected Reform Effects for the Long Run by the Reform Type

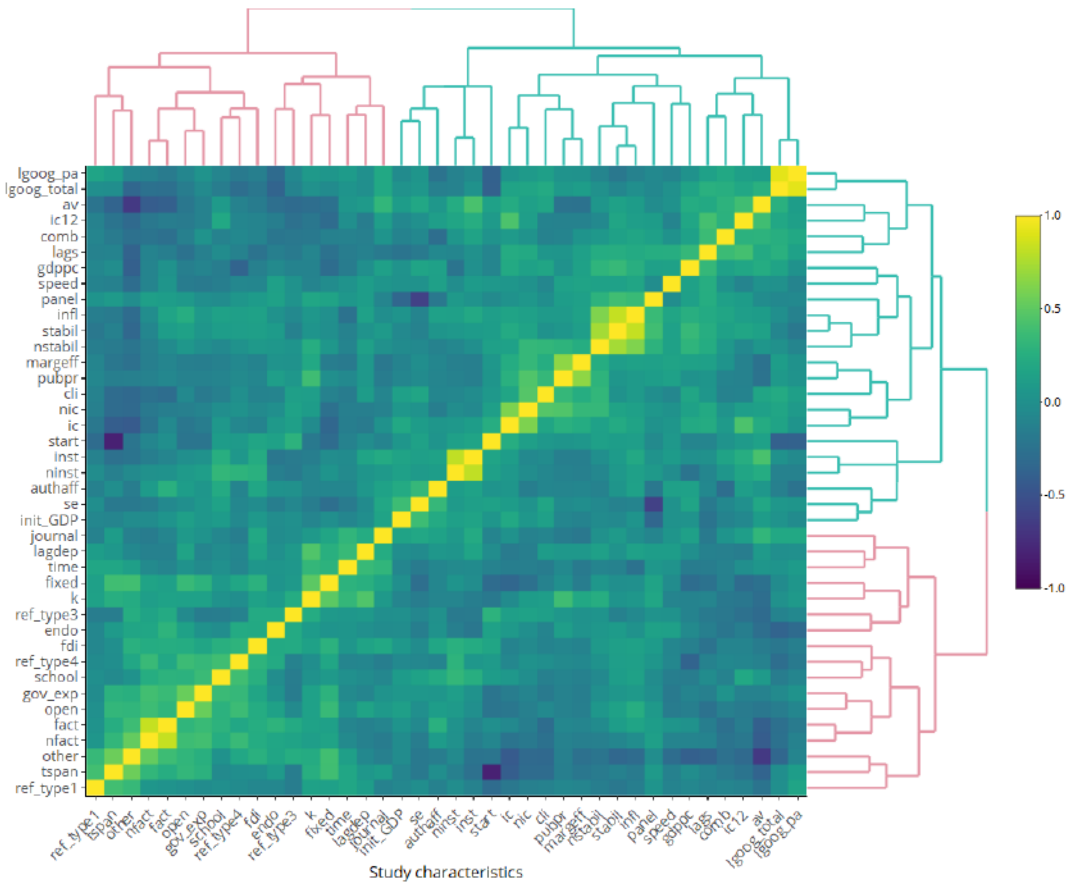


Notes: The vertical axis shows the degrees of freedom used for estimation of the reform effect in the primary study; the horizontal axis shows the long run reform effect represented by the partial correlation coefficient. The solid line denotes zero; the dashed lines denote the arithmetic means of the long run estimates by the reform type. The plot shows that the majority of estimates predicts a positive long-run impact of the reforms.

Appendix C

Results of BMA

Figure C.1: Correlation Matrix of Selected Study Characteristics



Notes: The figure shows Pearson correlations between selected study characteristics on a colour scale where yellow indicates positive correlation and purple negative one. Dendrograms on side display variable clusters with the horizontal axis (height) representing the dissimilarity between clusters.

Table C.1: Coefficient Estimates (Short Run)

Variable	PIP	Post. mean	Post. std. dev.	Cond. pos. sign
se	0.080	-0.036	0.177	0.092
Data and methods				
k	0.091	0.000	0.000	0.201
endo	0.068	0.000	0.008	0.618
fixed	0.955	0.115	0.042	1.000
start	0.425	0.000	0.000	1.000
tspan	1.000	0.006	0.001	1.000
gdppc	0.051	-0.001	0.015	0.428
Type of reform measure				
comb	1.000	-0.216	0.045	0.000
other	1.000	-0.281	0.045	0.000
av	0.256	0.019	0.039	0.960
cli	0.081	0.000	0.014	0.631
margeff	0.449	-0.049	0.064	0.000
ref_type1	0.283	-0.024	0.048	0.005
ref_type3	0.963	0.105	0.039	1.000
ref_type4	0.285	-0.049	0.093	0.000
Measure of dynamics				
lagdep	0.081	0.003	0.014	0.951
speed	0.896	-0.127	0.058	0.000
lags	1.000	-0.164	0.036	0.000
time	0.037	0.000	0.007	0.438
Specification characteristics				
ic	0.426	-0.037	0.050	0.000
ic12	0.021	-0.001	0.009	0.371
nic	0.810	-0.022	0.014	0.000
nstabil	0.094	-0.002	0.009	0.000
infl	0.429	-0.028	0.039	0.000
ninst	0.355	-0.014	0.021	0.000
school	0.993	-0.206	0.059	0.000
nfact	1.000	0.101	0.017	1.000
pubpr	0.200	-0.018	0.045	0.002
init_GDP	0.080	0.005	0.026	1.000
fdi	0.189	0.018	0.048	0.938
gov_exp	1.000	-0.213	0.047	0.000
open	0.688	0.089	0.075	1.000
Publication characteristics				
journal	0.140	-0.007	0.021	0.000
authaff	0.969	-0.113	0.039	0.000
lgoog_pa	0.013	0.000	0.001	0.325

Notes: PIP - Posterior inclusion probability; Post. mean - Weighted posterior mean; Post. std. dev. - Weighted posterior standard deviation; Cond. pos. sign - Conditional posterior sign.

Table C.2: Coefficient Estimates (Long Run)

Variable	PIP	Post. mean	Post. std. dev.	Cond. pos. sign
se	0.723	-0.366	0.273	0.000
Data and methods				
k	0.988	-0.005	0.001	0.000
panel	0.877	-0.090	0.047	0.000
endo	0.140	-0.004	0.012	0.000
fixed	0.754	0.066	0.045	1.000
start	0.980	0.000	0.000	1.000
tspan	0.027	0.000	0.000	0.171
Type of reform measure				
other	0.109	-0.007	0.027	0.049
av	1.000	0.194	0.028	1.000
cli	0.986	0.113	0.029	1.000
margeff	0.148	0.019	0.053	1.000
ref_type1	0.085	-0.008	0.030	0.000
ref_type3	0.089	-0.002	0.012	0.142
ref_type4	0.063	0.003	0.017	1.000
Measure of dynamics				
lagdep	0.096	-0.002	0.013	0.118
speed	1.000	-0.279	0.035	0.000
lags	0.224	-0.013	0.029	0.000
time	0.025	0.000	0.005	0.960
Specification characteristics				
ic	0.189	-0.007	0.023	0.136
ic12	0.668	-0.076	0.062	0.000
nic	0.374	-0.012	0.018	0.000
nstabil	0.065	-0.001	0.006	0.098
infl	0.714	0.048	0.037	1.000
ninst	0.062	-0.001	0.005	0.000
school	0.015	0.000	0.004	1.000
nfact	0.011	0.000	0.001	0.531
pubpr	0.055	-0.003	0.036	0.415
init_GDP	0.254	0.018	0.035	1.000
gov_exp	0.041	0.000	0.006	0.691
open	0.338	-0.027	0.044	0.000
Publication characteristics				
journal	0.058	0.001	0.007	0.977
authaff	0.052	-0.001	0.007	0.000
lgoog_pa	0.162	0.002	0.005	1.000

Notes: PIP - Posterior inclusion probability; Post. mean - Weighted posterior mean; Post. std. dev. - Weighted posterior standard deviation; Cond. pos. sign - Conditional posterior sign.

Appendix D

Diagnostics of BMA

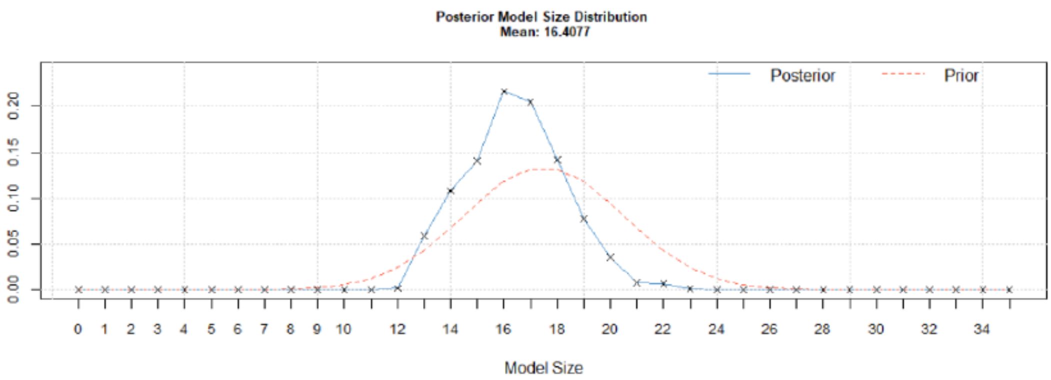
Table D.1: Summary of BMA Estimation (Short Run)

Mean no. regressors 16.4077	*Draws* 3000	*Burn-ins* 1000	*Time* 1.6361 secs
No. models visited 821	*Modelspace* 3.4×10^{10}	*% Visited* 2.4×10^{-6}	*% Topmodels* 100
Corr PMP 0.4718	*No. Obs.* 338	*Model Prior* uniform	*g-Prior* UIP
Shrinkage-Stats Av=0.9971			

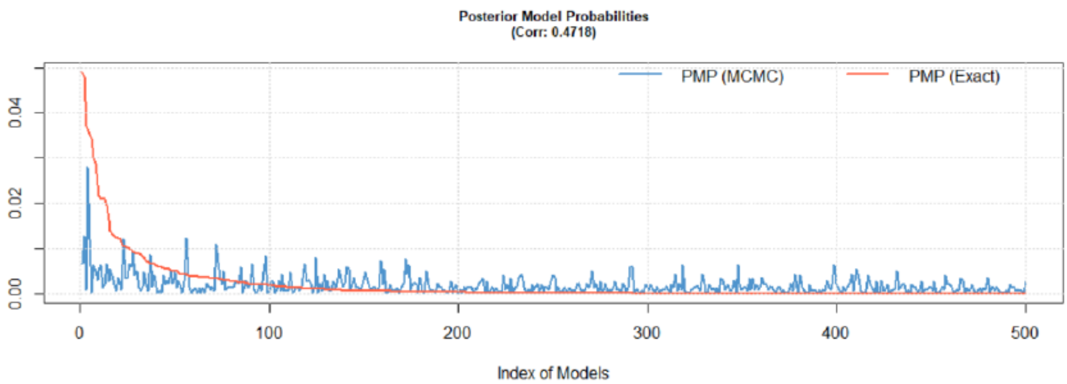
Table D.2: Summary of BMA Estimation (Long Run)

Mean no. regressors 11.3737	*Draws* 3000	*Burn-ins* 1000	*Time* 1.3063 secs
No. models visited 822	*Modelspace* 8.6×10^{9}	*% Visited* 9.6×10^{-6}	*% Topmodels* 100
Corr PMP 0.5482	*No. Obs.* 547	*Model Prior* uniform	*g-Prior* UIP
Shrinkage-Stats Av=0.9982			

Figure D.1: BMA Diagnostic Plots (Short Run)



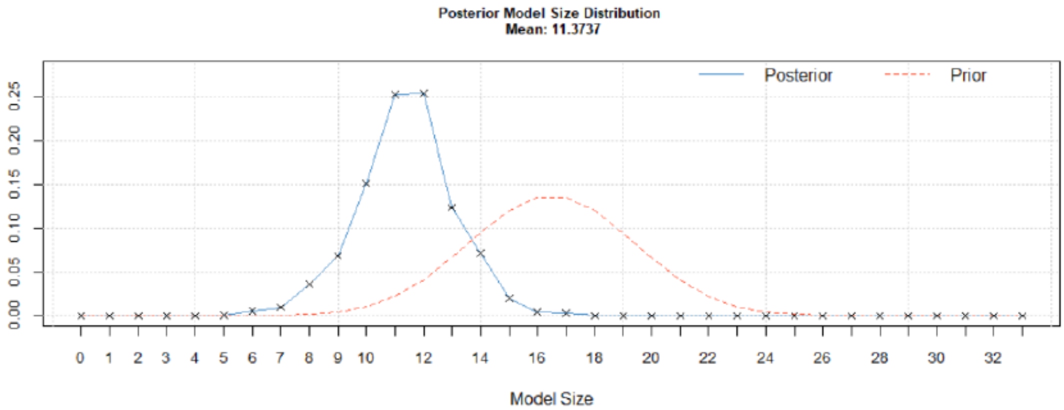
(a) BMA Model Size



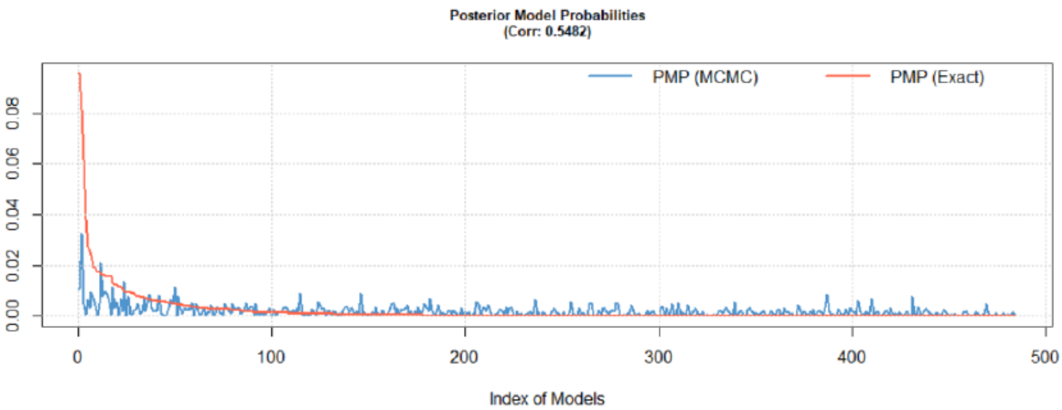
(b) Convergence Plot

Notes: Posterior model size distribution and model probabilities produced by the BMS package with uniform model priors on short-run reform effects.

Figure D.2: BMA Diagnostic Plots (Long Run)



(a) BMA Model Size



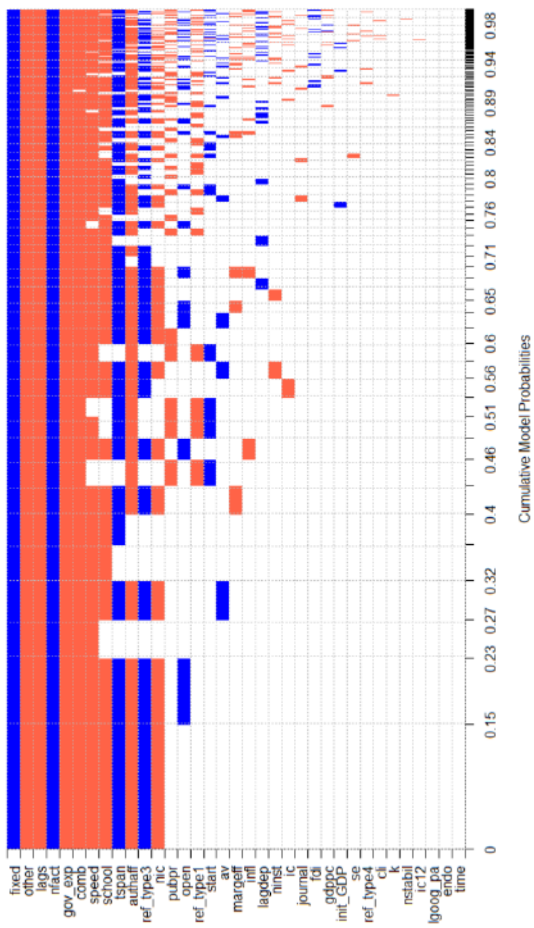
(b) Convergence Plot

Notes: Posterior model size distribution and model probabilities produced by the BMS package with uniform model priors on long-run reform effects.

Appendix E

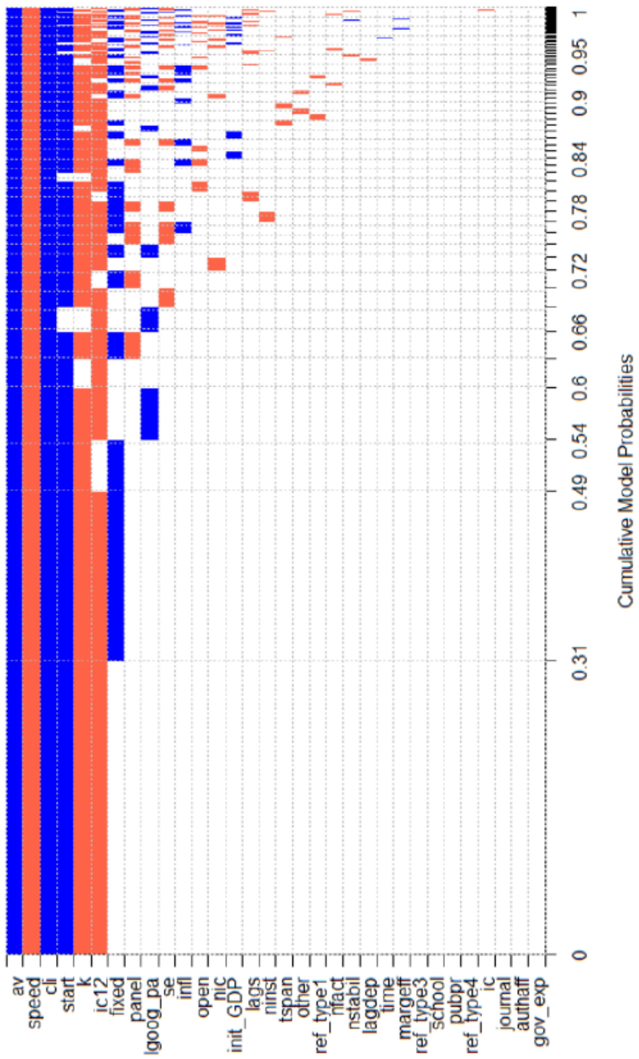
Robustness Check

Figure E.1: Bayesian Model Averaging, Model Inclusion (Short Run)



Notes: On the horizontal axis, the figure shows the best models, scaled by their PMPs. The explanatory variables are plotted on the vertical axis. Blue colour corresponds to a positive coefficient, red to a negative coefficient, and white to non-inclusion of the particular variable. Alternative priors are set to "random" and "BRIC".

Figure E.2: Bayesian Model Averaging, Model Inclusion (Long Run)



Notes: On the horizontal axis, the figure shows the best models, scaled by their PMPs. The explanatory variables are plotted on the vertical axis. Blue colour corresponds to a positive coefficient, red to a negative coefficient, and white to non-inclusion of the particular variable. Alternative priors are set to "random" and "BRIC".

Table E.1: Coefficient Estimates (Short Run)

Variable	PIP	Post. mean	Post. std. dev.	Cond. pos. sign
se	0.031	-0.020	0.134	0.065
Data and methods				
k	0.015	0.000	0.000	0.543
endo	0.022	0.000	0.004	0.788
fixed	1.000	0.157	0.037	1.000
start	0.275	0.000	0.000	0.994
tspan	0.872	0.004	0.002	1.000
gdppc	0.080	-0.005	0.024	0.124
Type of reform measure				
comb	0.994	-0.205	0.051	0.000
other	1.000	-0.246	0.052	0.000
av	0.088	0.006	0.021	1.000
cli	0.061	-0.002	0.015	0.489
margeff	0.151	-0.016	0.042	0.000
ref_type1	0.366	-0.052	0.076	0.000
ref_type3	0.538	0.055	0.056	1.000
ref_type4	0.052	-0.006	0.034	0.000
Measure of dynamics				
lagdep	0.163	0.009	0.024	1.000
speed	0.916	-0.132	0.057	0.000
lags	1.000	-0.174	0.039	0.000
time	0.003	0.000	0.002	0.000
Specification characteristics				
ic	0.054	-0.005	0.021	0.000
ic12	0.016	-0.001	0.008	0.149
nic	0.447	-0.012	0.014	0.000
nstabil	0.049	-0.001	0.005	0.054
infl	0.215	-0.013	0.028	0.000
ninst	0.179	-0.007	0.016	0.000
school	0.860	-0.162	0.085	0.000
nfact	1.000	0.095	0.020	1.000
pubpr	0.424	-0.056	0.075	0.000
init_GDP	0.016	0.002	0.014	1.000
fdi	0.203	0.024	0.059	0.962
gov_exp	0.995	-0.188	0.046	0.000
open	0.305	0.037	0.063	1.000
Publication characteristics				
journal	0.056	-0.002	0.010	0.006
authaff	0.818	-0.085	0.051	0.000
lgoog_pa	0.034	0.000	0.002	0.265

Notes: PIP - Posterior inclusion probability; Post. mean - Weighted posterior mean; Post. std. dev. - Weighted posterior standard deviation; Cond. pos. sign - Conditional posterior sign.

Table E.2: Coefficient Estimates (Long Run)

Variable	PIP	Post. mean	Post. std. dev.	Cond. pos. sign
se	0.158	-0.058	0.166	0.110
Data and methods				
k	0.815	-0.004	0.002	0.000
panel	0.256	-0.021	0.040	0.000
endo	0.000	0.000	0.000	NA
fixed	0.472	0.039	0.047	1.000
start	0.892	0.000	0.000	1.000
tspan	0.032	0.000	0.000	0.000
Type of reform measure				
other	0.067	-0.003	0.017	0.124
av	1.000	0.195	0.027	1.000
cli	0.984	0.110	0.030	1.000
margeff	0.017	0.002	0.017	1.000
ref_type1	0.021	-0.002	0.017	0.000
ref_type3	0.043	-0.002	0.009	0.000
ref_type4	0.003	0.000	0.004	1.000
Measure of dynamics				
lagdep	0.035	-0.002	0.012	0.000
speed	1.000	-0.273	0.035	0.000
lags	0.072	-0.005	0.020	0.000
time	0.043	0.000	0.006	0.805
Specification characteristics				
ic	0.010	0.000	0.004	0.000
ic12	0.822	-0.110	0.061	0.000
nic	0.065	-0.002	0.007	0.000
nstabil	0.043	0.000	0.004	0.223
infl	0.163	0.011	0.027	1.000
ninst	0.054	-0.001	0.005	0.000
school	0.031	0.001	0.007	1.000
nfact	0.052	0.000	0.003	0.045
pubpr	0.016	-0.002	0.022	0.000
init_GDP	0.061	0.003	0.015	1.000
gov_exp	0.002	0.000	0.001	0.000
open	0.152	-0.012	0.031	0.000
Publication characteristics				
journal	0.006	0.000	0.002	1.000
authaff	0.006	0.000	0.002	0.588
lgoog_pa	0.230	0.004	0.008	1.000

Notes: PIP - Posterior inclusion probability; Post. mean - Weighted posterior mean; Post. std. dev. - Weighted posterior standard deviation; Cond. pos. sign - Conditional posterior sign.