CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

How long does it take until the positive effects of structural reforms do materialize?

Bachelor thesis

Author: Jan Hanzal

Study program: Ekonomie a finance

Supervisor: PhDr. Jaromír Baxa, PhD

Year of defense: 2020

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 5, 2020

Jan Hanzal

Abstract

This thesis studies the topic of the effects of structural reforms over time. The research performed until now has suggested that the dynamic effects of reforms are clear, with neutral effects or costs in the short term, and important benefits in the medium to long term. In order to verify this seemingly settled view, this thesis tests the robustness of a well-established research paper on this topic by using an extended dataset and performing modifications to the measurement of several variables in the underlying model. Our results do not confirm the usual hypotheses, showing effects that are mostly not statistically or practically significant.

Keywords Structural reforms, Local projections, Reform measurement, Dynamic effects

Title How long does it take until the positive effects of structural reforms do materialize?

Abstrakt

Tato práce se zabývá tím, jak se vyvíjí efekty strukturálních reforem po jejich uskutečnění. Dosavadní výzkum lze shrnout poměrně jasně, jelikož většina vědeckých prací se shoduje na tom, že strukturální reformy mají negativní nebo neutrální vliv na ekonomiku v krátkém období, a značně pozitivní vliv ve střednědobém až dlouhodobém horizontu. Abychom otestovali robustnost těchto výsledků, zkoumali jsme, jak ovlivní výsledky jedné z prací, která se tímto tématem zabývá, rozšíření datasetu a využití jiných metod pro měření určitých proměnných. Naše výsledky jsou značně odlišné. Efekty strukturálních reforem v naší práci totiž většinou nejsou statisticky významné, případně jsou relativně malé.

Klíčová slova Strukturální reformy, Local projections, Měření reforem, Dynamické efekty

Název práce Jak dlouho trvá, než se projeví pozitivní efekty strukturálních reforem?

Acknowledgments

The author is grateful to PhDr. Jaromír Baxa, PhD for his guidance and valuable advice during the writing of this thesis.

Typeset in FSV LaTeX template with great thanks to doc. PhDr. Zuzana Havránková Ph.D. and doc. PhDr. Tomáš Havránek Ph.D.

Bibliographic Record

Hanzal, Jan: *How long does it take until the positive effects of structural reforms do materialize?*. Bachelor thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 74. Advisor: PhDr. Jaromír Baxa, PhD

Contents

List of Tables

List of Figures

Acronyms

EPL Employment Protection Legislation

ETCR Regulation in Energy, Transport and Communications

GAM Generalized Additive Model

HP filter Hodrick-Prescott filter

IMF International Monetary Fund

OECD Organization for Economic Cooperation and Development

OLS Ordinary Least Squares

PMR Product Market Regulation

VAR Vector Autoregression

Bachelor's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author's name and surname: Jan Hanzal

E-mail: 23783855@fsv.cuni.cz

Phone: +420 773 997 108

Supervisor's name: PhDr. Jaromír Baxa, PhD.

Supervisor's email: jaromir.baxa@fsv.cuni.cz

Notes: Please enter the information from the proposal to the Student Information System (SIS) and submit the proposal signed by yourself and by the supervisor to the Academic Director ("garant") of the undergraduate program.

Proposed Topic:

How long does it take until the positive effects of structural reforms do materialize?

Preliminary scope of work:

Research question and motivation

The aim of this thesis will be to empirically assess the amount of time it takes for structural reforms to have a positive impact on the economy.

The European Central Bank[1] defines structural reforms as "measures that change the fabric of an economy" and that allow for balanced growth of that economy.

Those are measures that should promote economic growth, at least in the long term (de Bandt & Vigna, 2008). However, the question stands whether there are any costs to the structural policies, and if so, how long it takes before the benefits outweigh the costs. The answer to this question is important to determine since it is helpful in assessing, for example, whether to use these policies to mitigate recessions, which are obviously short-term phenomena. It has been found that reforms are more often undertaken when the economy is not doing well (Dias et al., 2017). But what is the short-term effect of these policies? Opinions on this differ, and we often find they are in direct contradiction with each other.

OECD (Bouis et al., 2012) states that structural reforms are generally beneficial even in the short term. They find that structural reforms can entail costs only in very specific cases. In the language of our thesis, this would mean that the positive effects materialize almost immediately.

A very different view is presented by De Grauwe, & Ji (2016) who studied the effect of structural reforms with regard to asymmetric shocks in the eurozone. They infer that structural reforms introducing more flexibility into the economy help in case of permanent asymmetric shocks. However, with short term shocks in the form of busts, the structural reforms only aggravate the situation. Furthermore, they found only a small impact of structural reforms on long-term growth in the eurozone.

Eggertsson, Ferrero and Raffo (2014) conclude from their simulations that structural reforms, while efficient in the long term, can have serious short-term consequences for the economy if the monetary policy is at the zero lower bound.

Babecký and Havránek (2013) performed a meta-analysis of papers on the relationship between structural reforms and economic growth. They conclude that the studies confirm that reforms generate costs in the short-term of about one-year and then generate benefits in the following periods.

The variation of the results depends largely on the methods used, the type of reforms chosen, the initial conditions, as well as the data used (especially the choice of structural reforms indicators) (Babecký & Campos, 2011). I will therefore focus on these issues when performing the analysis.

Contribution
This thesis will empirically investigate the question stated above. Rather than concentrating on the magnitude of the impact of structural reforms as most studies do, it will focus on estimating the time before the effects take place. Before performing any regressions, I will analyse the definitions of various reforms, define interactions between the reforms (as proposed by Babecký & Campos, 2011), as well as carefully decide which data sources (of the reform indicators) to use and why, and consider other factors, taking into account the research that has already been done. The proposed reasearch is especially relevant because of the apparent lack of this kind of research in recent years when data collected during the Great Recession and in the immediate aftermath has become available.

Methodology
The appropriate methodology will be chosen after a careful literature review. I plan to exploit conventional macroeconomic data that are widely available.

Outline

1. Introduction
    - What are structural reforms
    - Why it is important to study how long it takes for them to materialize
2. Literature review
    - Various definitions of reforms, their classification
    - What influences their success/failure
    - Overview of data sources
    - Overview of existing models
3. Methodology
    - What methodology & data was chosen, detailed description of the model
    - How the analysis was performed
4. Results
    - Answer to the research question
5. Conclusion
    - What can be drawn from the results, potential policy implications
    - What could be studied in the future

List of academic literature:

Bibliography

Babecký, J., & Campos, N. F. (2011). Does reform work? An econometric survey of the reform-growth puzzle. *Journal of Comparative Economics*, *39*(2), 140–158. https://doi.org/10.1016/j.jce.2010.11.001

Babecký, J., & Havránek, T. (2013). Structural Reforms and Economic Growth: A Meta-Analysis. *CNB Working Paper Series*, (8).

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

de Bandt, O., & Vigna, O. (2008). The macroeconomic impact of structural reforms. *Banque de France Bulletin, Quarterly Selection of Articles*, (Spring 2008).

De Grauwe, P., & Ji, Y. (2016). Flexibility Versus Stability: A Difficult Tradeoff in the Eurozone. In *Credit and Capital Markets – Kredit und Kapital* (Vol. 49). https://doi.org/10.3790/ccm.49.3.375

Dias, A., Silva, D., & Givone, A. (2017). When do countries implement structural reforms ? *ECB Working Paper*, (2078).

Eggertsson, G., Ferrero, A., & Raffo, A. (2014). Can structural reforms help Europe? *Journal of Monetary Economics*, *61*(1), 2–22. https://doi.org/10.1016/j.jmoneco.2013.11.006

Chapter 1

Introduction

Structural reforms are fundamental changes in an economy which are undertaken with a view of future benefits. Thanks to this wide definition, they can take on many forms, for example: product market reforms, labour market reforms, financial market reforms, trade reforms, tax reforms and others. Intuitively, one can assume that since the reforms improve the structure of an economy and thus its competitiveness, they result in long term benefits. According to de Bandt *et al.* (2008), this is because structural reforms increase competition and make the economy more robust to shocks.

A question bids itself however: what are the actual, empirical effects of these reforms and how long does it take for positive effects to appear? Since the character of the reforms does not lie in a short-term boost of the economy, but in a rather permanent and structural change, it is necessary to observe their effects over some period of time, rather than just look at the immediate effects. This and the fact that reforms are not directly measurable are the main issues when choosing empirical methodology to study their effects. The existing literature has therefore adopted various ways to measure reform. Furthermore, a specific econometric model has been shown to be particularly suited to the task at hand: the local projections model. Despite the various smaller discrepancies in the results, there seems to be a consensus in the literature that reforms have positive effects in the long run and neutral or negative effects in the short run. However, the robustness of the results is not quite certain. There has been a limited number of studies, such as that of De Grauwe *et al.* (2016) which show that reforms in general do not have significant effects. Moreover, in a meta-analysis concentrated on the effects of reform in transition economies, Babecky & Campos (2011) show that the results tend to be very sensitive to various model specifications. Furthermore, in a similar meta-analysis, Babecky & Havranek (2013) show that a publication bias exists in reporting the short-term effects of structural reforms in transition economies.

It is important to note furthermore that a large portion of the research is not undertaken by primarily academic institutions, but rather by various international organisations, public or private, such as the International Monetary Fund, the Organisation of Economic Co-operation and Development or the IZA Institute of Labor Economics. Therefore the publication of results might be influenced by various policy aims. Indeed, the results of the research on structural reforms are relevant to policy making: the estimated effects of reforms might indicate when the reform should be undertaken, whether during crisis or during a period of growth, whether it needs to be supported by other policy, or whether it is a relevant policy instrument at all. Therefore it is important to check whether the published results are robust to various changes in the model specification or in the data used, since important policy decisions could be taken on the basis of these results.

To test the robustness of the research, we decide to go down the path of replication and subsequent extension of a well-established paper on the dynamic effects of reforms written by Bordon *et al.* (2016). Hamermesh (2007) describes two basic categories of replication in economic research: *pure replication* which aims at replicating a given paper using the very same approach and data, and *scientific replication*, which uses different data or an approach that is similar, but not identical to that of the original author. This thesis concentrates on the latter.

Inspired by Bordon *et al.* (2016), we extend the analysis in a number of ways. We focus first and foremost on extending the dataset in the country and time dimensions. The interest is to cover the topic without imposing unjustified restrictions on the country sample. Furthermore, since the structure of the world's economies seems to have been influenced by the financial crisis of 2008, the extension in the time dimension allows to study whether this has also affected the effects of reforms. Moreover, the time extension is also beneficial from a methodological perspective, as it will be explained later. This extension is also followed by further robustness checks - namely a restriction of the dataset in the manner of Bordon *et al.* (2016). Moreover, to our knowledge, no study has concentrated on the influence of output gap measurement on the measurement of reform effects. The output gap is however, as we will see, an important control variable and this is therefore also addressed in this thesis.

Lastly, since reform measurement is an important question in the research on structural reforms, we replace the reform data with a narrative dataset as a further robustness check.

Our results suggest that the established view on the effects of structural reforms should perhaps be taken with a grain of scepticism. Most of our results do not show statistical significance. Those that do, often show smaller effects than estimated e.g. in the paper by Bordon *et al.* (2016). We also present evidence suggesting possible negative effects of product market reforms even in the longer term.

The rest of this thesis is organized as follows. Chapter 2 comprises of the literature review, describing various approaches to reform measurement and to empirical methodology. Chapter 3 then concentrates on the data and methodology used in this paper and describes the data extension vis-a-vis Bordon *et al.* (2016). Chapter 4 shows the results. This is then followed by the robustness checks in chapter 5. Finally, we conclude by commenting on the results as a whole and on their possible implications.

Chapter 2

Literature review

In this chapter, we will introduce some of the specific issues connected with estimating the dynamic effects of structural reforms that are described in the literature. First, the thesis introduces the problem of structural reforms measurement. Then, it concentrates on how the relevant literature uses econometric methodology to estimate the dynamic effects.

2.1 Measuring structural reforms

When performing any econometric analysis of structural reforms, it is essential to decide how a reform will be measured, such that variables indicating that a reform has occurred can be used in a regression. Let us first look at the various macroeconomic and other indicators used in the literature on the effects of structural reforms. Then, we will explore how the literature exploits those indicators as variables in an econometric model.

2.1.1 Reform and regulation indicators

There are quite a few indicators of labour reforms used in the literature. One of the most common measures are the OECD Employment Protection Legislation (EPL) indicators. These indices summarize the degree to which workers with regular contracts are protected from dismissals, the protection of workers from collective dismissals and also the amount of regulation of temporary employment contracts (with respect to regular contracts) (OECD 2013). Among the papers using these indicators is a text on the influence of the business cycle and macroeconomic policies on the effects of structural reforms written by Bordon *et al.* (2016). The authors of this study appreciate that the EPL indices are comprehensive, but also note that the problem with them is that there are some reforms that they do not capture quite accurately. Most other papers, studying labour market reforms, that are mentioned here have also used these indices, among them e.g. Bouis *et al.* (2012b). The EPL indices are available mostly only for OECD member countries. An index which is available for a wider array of countries is the World Bank mandated cost of worker dismissal (de Almeida & Balasundharam 2018). However, by using this index, the authors take a narrower view of employment market reforms. The availability of comprehensive indicator data seems to be one of the reasons why most studies focus on developed countries.

Another important indicator in the literature is the amount the governments spend on active labour market policies, which are deemed to reduce rigidities in the labour market, allowing for a more effective allocation of workers (Bouis *et al.* 2012b; Egert 2017). Bouis *et al.* (2012b) describe the short-term effects of structural reforms. Among the more unorthodox indicators, the authors also use measures of unemployment benefits, labour taxes, proportion of direct taxes on tax revenue, minimum retirement age, and difference between workers covered by a collective agreement and number of union members. Another very similar study to that of Bouis *et al.* (2012b), written by an almost identical group of authors and concentrating solely on labour reforms, Bouis *et al.* (2012a), uses an OECD indicator of unemployment benefit replacement rates. Therefore, rather than concentrating on reforms undertaken in the domain of employment protection legislation, the authors focus solely on unemployment benefit reforms. As we can see, there is a wide variety of indices that can be used for estimating the effects of labour market reforms. When choosing an index, either general or specific, there seems to be a trade-off between generality and accuracy.

When it comes to product market reforms, the resources seem to be scarcer. The OECD publishes the Product Market Regulation (PMR) indicators which summarize the degree to which regulation impedes competition and firm entry for essentially the whole country's economy (OECD 2018). However, new data is published only once every five years, therefore without further modifications, it is unusable as yearly data. In a study on the impact of regulation on unemployment, Piton & Rycx (2018) do use the PMR indicator, however they estimate a static model, where it is not necessary to have data for each country in the form of a continuous yearly time series. In their firm-level analysis on the short run effects of product market reforms, Gal & Hijzen (2016) create a yearly database based on the PMR indicator, filling in data for the intermediate years by relying on textual information on structural reforms collected by the IMF. Unfortunately, the authors do not seem to share their database, nor the IMF data source. Most often however, the literature uses the OECD Regulation in Energy, Transport and Communications (ETCR) indicators as proxy variables for the more widely defined PMR indicators, since the data for these is available yearly (for example, Tola & Waelti (2018); Bordon *et al.* (2016); Bouis *et al.* (2012b) use the ETCR indicators). The ETCR data covers only the network industries of: electricity and gas; air, rail and road transport; post and telecom. Similarly to the PMR indicators, it contains information on entry regulation, public ownership, vertical integration and market structure of the industries (Koske *et al.* 2015).

The sources tend to be very varied for other types of reform. As for financial reform, a paper on the effect of structural reforms on productivity growth in developing countries by Dabla-Norris *et al.* (2016) uses indicators of financial regulation and reform compiled by Abiad *et al.* (2008). The data includes information on credit and interest controls, excessive reserve requirements, capital account restrictions and other measures of regulation restrictiveness. These indices are also used by Christiansen *et al.* (2009). As for trade reforms measures, these are largely based on the size of the tariff rates and other barriers to trade (Christiansen *et al.* 2009).

2.1.2 Usage of indicators as variables

As for the application of the regulation and reform indicators to an econometric model, possibly the most straightforward approach is to directly use the indicators as independent variables. Bassanini (2015) uses this approach to study the effects of product market reforms. As the variable indicating a reform, the author uses the continuous change in the ETCR indicator. By doing this, the author takes fully into account the size of the change of the indicator. Another study (Piton & Rycx 2018) uses the OECD Product market regulation and Employment protection legislation indices directly in a regression, without performing any transformation except for demeaning (for purposes of estimating a fixed effects regression). Egert (2017) also uses the regulation indicators directly to study the effects of regulation on multi-factor productivity. This paper however does not focus on the impact of structural reforms *per se*, but rather on the impact of the level of *regulation*. When using this approach for *reforms*, one has to question whether any small change in a given indicator really constitutes a reform, which is largely understood as being more of a larger shock rather than a small change.

Another possible approach, which actually seems to be the most frequently used one in the literature, is to use dummy variables. Generally, this means that when the change in a given indicator falls below a certain threshold, the dummy variable indicating whether a reform has happened is equal to one. Bordon *et al.* (2016) use the threshold of two standard deviations of the change in the ETCR indicator, and one standard deviation of the change in the EPL indicator. The author justifies this by the fact that labour market reforms are much less numerous than product market reforms in the sample (given the same threshold). A similar approach is outlined in a paper on growth and structural reforms by Christiansen *et al.* (2009). The researchers put the reform dummy equal to one when there is *any* change in a given indicator (the reform is either a positive reform or a reform reversal, depending on the sign of the change), while they also define a dummy for *large reforms* when the change is greater than one standard deviation. A modification of this method is to multiply the dummy variable by the size of the indicator change. The variable is then equal to the size of the change only if a reform is detected (Bouis *et al.* (2012a); Bouis *et al.* (2012b)). The authors of the studies using this approach emphasize that this allows to control for the size of the reform, as opposed to using a plain dummy variable. The threshold they use is two standard deviations. The most obvious issue with the dummy variable approach is that the threshold is arbitrary, which the authors admit (Bouis *et al.* (2012a); Bouis *et al.* (2012b)). The solution they propose is to do a sensitivity analysis using different thresholds. Dabla-Norris *et al.* (2016) supplement the approach based on standard deviations with a so-called "up-break" approach, which requires that the change in the indicator also lasts for a certain period of time (i.e. is not immediately reversed), based on the methodology outlined in Berg *et al.* (2012).

Yet another different way to use information on structural reforms is through narrative databases. These have not been used as much in the literature due to the fact that they were not available until recently. Narrative databases are constructed such that they use several data sources to decide whether a reform occurred or not. It is the author of the database who decides whether the information is sufficient to indicate whether a reform happened (Romer & Romer (1989), here the authors introduce this approach to identify monetary disturbances). The authors of an IMF (2016) paper, who use a narrative dataset for estimating the effects of structural reforms, claim that the advantages of this approach are the precision of the identification of the reform (a mere change in an indicator is not enough) and the fact that the data can be constructed even for reforms for which no relevant indicators exist. A very recent study by Alesina *et al.* (2020) uses a self-created database of structural reforms, covering a wide variety of them, as well as a wide range of countries, to estimate the effects of structural reforms on elections. A publicly available dataset covering ETCR and EPL reforms has been compiled by Duval *et al.* (2018). This has been used in a paper concentrating on the effects of employment protection deregulation depending on the state of the economy (Duval *et al.* 2017) as well as in yet another study on the dynamic effects of structural reforms, which also takes into account the state of the economy and the macroeconomic policy reactions (Duval & Furceri 2018).

2.2 Econometric methodology

Static models. One option to estimate the dynamic effects of structural reforms is to use static panel data methods. A static model, as opposed to a dynamic panel data model, does not contain lags of the variables. Piton & Rycx (2018) use a static fixed effects model to estimate the effects of EPL and PMR deregulation on the unemployment rate. In order to differentiate between the short run and long run effects of reforms, the authors decided to perform two regressions: one regression excluding the countries with large recent reforms (Greece, Italy, Portugal, Spain) and a different one with the whole sample. They concluded that, excluding the four countries, the effect of an EPL deregulation was positive, and when the regression was performed on the full sample, the effect was negative. Thus the authors inferred that the effect is negative in the short run and positive in the long run.

Christiansen *et al.* (2009) use a static model to infer the general effect of structural reforms on growth variables. However, they note that when using this model, it is impossible to differentiate between short run and long run effects of structural reforms. Bassanini (2015) even notes that using a static model with regard to estimating only the immediate effect of a reform on employment would result in a biased estimate, since there might be a common confounding factor influencing both the dependent variable and the reform. To estimate short term effects only, the author suggests lagging the reform variable by one period (without necessarily using the present period value in the equation as well). To avoid endogeneity, it is also possible to use the instrumental variable approach, which is done by De Grauwe *et al.* (2016), who estimate the effects of reforms on growth in a static model (they also include other measures that are supposed to increase an economy's flexibility). The authors find that reforms in general do not have significant effects on economic growth.

Standard dynamic panel data models. Dynamic models are perhaps a little more suited for the task at hand, since they explicitly allow to lag the reform variables. When studying the impact of reforms on multi-factor productivity, Egert (2017) uses one lag and one lead of the dependent variables in his first-differences estimation. However, it is important to note that the paper does not concentrate on the dynamic effects of reforms and the leads and lags are included to avoid a possible source of endogeneity of the independent variables. This shows another degree of usefulness of the dynamic estimation method versus the static one.

Some of the authors using the aforementioned static models also estimate dynamic ones. Christiansen *et al.* (2009) include six periods of lags of the reform variable, thus allowing them to study the effects of those lags on the dependent variable in the current period. Not only that, but the authors also add leads of the reform variable, so that they can control for possible reverse causality, i.e. to control for cases when good economic conditions lead to reforms in the future. To check for the correct number of lags, the authors also perform robustness checks by adding up to 9 lags. Bassanini (2015) has a very similar approach, but uses the Bayesian Information Criterion and the Akaike Information Criterion to decide on the number of lags to be used.

Local projections models. The task at hand is essentially to estimate a so-called impulse-response function. The goal is to estimate what effect a reform has at certain time horizons in the future, i.e. what response (the effect on the dependent variable) an impulse (a reform) produces. Traditionally, vector autoregression has been performed to estimate the impulse-response function. However, a much simpler way of estimating the impulse-response function is to use the local projections presented by Jorda (2005), which entails estimating the following equation:

$$y_{t+s} = \alpha_s + \beta_{1,s} y_{t-1} + \beta_{2,s} y_{t-2} + \ldots + \beta_{p,s} y_{t-p} + u_{t+s}, \tag{2.1}$$

where y_t is the variable of interest (which can experience a shock), $t = 0, \ldots, T$ indicates the time dimension and the equation is estimated for each $s = 0, \ldots, h$, where h is the maximum horizon of the impulse response function that a researcher wants to estimate. According to the author, the method has a number of advantages over VAR, including simplicity of estimation (standard OLS estimation can be used) and robustness to misspecification. Furthermore, a paper by Plagborg-Møller & Wolf (2019) proves that the impulse-response function estimated by both methods is identical. The specification in the literature on structural reforms tends to be slightly different in that it estimates the impulse-response from panel data, with control variables and using fixed effects, similarly to Teulings & Zubanov (2010).

The local projections method has been used by many papers on the subject. By estimating the local projections, Bouis *et al.* (2012b) find that structural reforms have largely positive and significant effects on unemployment, with costs being felt only in a limited number of cases and only in depressed economies. Bouis *et al.* (2012a) find similar results, noting that negative effects show only when an economy is not doing well. Bordon *et al.* (2016) estimate that the effect of product market reforms on employment is almost only positive, while labour market reforms start to have a positive and significant effect in the fourth and fifth years after a reform occurs. Dabla-Norris *et al.* (2016) estimate the effect of structural reforms on total factor productivity in non-developed economies. They also note that there seem to be no important costs, however, the reforms take time before they materialize. Furthermore, the effect varies to a significant degree across different types of reforms. Using a a sectoral dataset Duval *et al.* (2017) find that labour market reforms have significant positive effects when the economy is doing well and negative effects in a recession. Duval & Furceri (2018) show that active monetary and fiscal policy can improve the effects of reforms. To sum up, most studies therefore show that reforms have a largely positive and significant effect in the medium and long term. There is also some evidence for short term costs.

Chapter 3

Empirical estimation

In this chapter, the thesis will estimate the dynamic effects of reforms. The equations that are estimated are largely based on the first part of a paper by Bordon *et al.* (2016) called: "When Do Structural Reforms Work? On the Role of the Business Cycle and Macroeconomic Policies". The aim of this thesis is to test the robustness of their model to new data and different measures of certain variables. The paper is particularly interesting for two reasons: it uses the local projection method, which, as it can be seen in the literature review, is the primary method used in economic research to estimate the dynamic effects of structural reforms. Secondly, it uses quite a similar specification to many other papers on the subject which have been mentioned in the literature review, e.g. Bouis *et al.* (2012b), Bouis *et al.* (2012a), IMF (2016), Tola & Waelti (2018). Therefore the results on the robustness of the method used in the paper by Bordon *et al.* (2016) can suggest similar outcomes for a wide array of papers on the topic.

The primary difference between the original paper and our approach is the addition of new data. Furthermore, we use different methods to estimate several variables, primarily the output gap. We also add the GDP as a further dependent variable. Following that, as a robustness check, data is restricted to match that of Bordon *et al.* (2016). Lastly, further robustness checks are performed to test whether the results are robust to different approaches of reform and output gap measurement.

In the following chapters and sections, first there will be a description of the methodology and of the data. Then, the results follow. Lastly, robustness checks are performed.

3.1 Methodology

In terms of methodology, the thesis follows Bordon *et al.* (2016) [1] , using the local projections method proposed by Jorda (2005). This entails estimating the following equations through fixed effects regression. For each $h = 0, 1, 2, 3, 4, 5$, we estimate a separate equation, with the dependent variable $y_{i,t}$ being either employment (as in Bordon *et al.* (2016)) or the logarithm of the GDP:

$$\begin{aligned} y_{i,t+h} - y_{i,t-1} =& \beta_{0,h} + \beta_{1,h} R_{i,t} + \beta_{2,h} BC_{i,t} + \beta_{3,h} OG_{i,t} + \beta_{4,h} OG_{i,t-1} + \\ &+ \sum_{k=1}^{3} \beta_{k+4,h}(y_{i,t-k} - y_{i,t-k-1}) + \beta_{8,h} OLC_{i,t} + CFE_i + TFE_t + u_{i,t+h} \end{aligned} \tag{3.1}$$

where:

- $i = 1, ..., N, t = 1, ..., T$ are the country and year indices, respectively,
- $y_{i,t+h} - y_{i,t-1}$ is the difference in the dependent variable (employment in the original study) between time $t + h$ and $t - 1$,
- $R_{i,t}$ is the reform shock,
- $BC_{i,t}$ is the banking crisis dummy,
- $OG_{i,t}$ is the output gap,
- $y_{i,t-k} - y_{i,t-k-1}, k = 1, 2, 3$ are the lags of the difference in the dependent variable,
- $OLC_{i,t}$ is the output loss during crisis,
- CFE_i and TFE_t are the country and time fixed effects, respectively,
- $u_{i,t+h}$ is the disturbance.

The focus is on the coefficients $\beta_{1,h}$ for each $h = 0, ..., 5$. They estimate the effect of the reform on a given "lead" of the dependent variable, which is in line with a basic regression analysis. Their sequence then provides the estimate

[1] Bordon *et al.* (2016) estimate the following equation:
$e_{i,t+h} = \theta_h R_{i,t} + \psi_h(L)e_{i,t+h-1} + X^T_{i,t-1}\Gamma_h + fixed\_effects + \epsilon_{i,t+h}$,
where $e_{i,t+h} = E_{i,t+h} - E_{i,t-1}$, $E_{i,t}$ is the employment rate and $X^T_{i,t-1}$ the vector of control variables. Our equation does not include $\psi_h(L)e_{i,t+h-1}$. For justification, see the Appendix A.

of the impulse-response function, while the standard errors of the parameter quantify the uncertainty of this estimate.

As for the control variables, they are meant to capture the specific economic conditions, which might be correlated with the reforms taking place, as well as with the dependent variable. According to Bordon *et al.* (2016) the cyclical conditions influence the probability of a reform being implemented, while obviously influencing unemployment as well. Therefore, there is the danger of mistaking an upturn or a downturn in the economic cycle with the effect of a reform. For this reason the authors of the original paper included the output gap and its lagged value in the regression. In addition, to control for actual banking crises, which might be related with the two variables as well, they introduced a banking crisis dummy. The cumulative output loss also controls for the severity of the crisis. As opposed to the output gap, this number represents the total losses of the crisis, not just that of the current year. It might be important to differentiate between the two since a crisis, which endures long but might induce only small losses in a given year, would result in very different magnitudes of the two relevant variables. To control for the level of employment in previous years, which also might influence the implementation of a reform later on (high unemployment in the recent past could incite the government to take action, which might however take some time for reasons such as the legislative process), three lags in the first difference in employment are used.

Equation (3.1) can either be estimated using OLS, or using Fixed Effects (two-ways) estimation if the country and time dummies are left out. The country fixed effects are justified by the fact that a large cross-country heterogeneity can be found in the case of reform implementation and economic conditions. Leaving these effects out of the equation would result in: $u_{i,t+h} = CFE_i + \varepsilon_{i,t+h}$, where $\mathbb{E}(CFE_i|X) \neq 0$ (X being the matrix of independent variables), therefore the disturbance would be correlated with the regressors, resulting in biased estimation (Wooldridge 2002). The same logic can be applied to time fixed effects. An example of a correlated country individual variable could be the specific institutions or political culture in a given country that affect the implementation of reforms as well as the economic conditions (employment). An example of a common time variable could be various global economic shocks that also affect both the dependent variable and the reform variable and which have approximately the same effect across countries, such as oil crises or global financial crises.

Adding country and year dummies eliminates these fixed effects by explicitly controlling for these variables. The other possibility is to subtract the means for both each country and each time period. In the case of the country fixed effects, Wooldridge (2002) shows the elimination of the bias in the following manner. First, average the equation across the time dimension:

$$\bar{y}_i = \bar{\boldsymbol{x}}_{\boldsymbol{i}}\boldsymbol{\beta} + CFE_i + \bar{u}_i \tag{3.2}$$

where $\bar{\boldsymbol{x}}_{\boldsymbol{i}}\boldsymbol{\beta}$ is the dot product of all the time-demeaned variables and their respective parameters. Then subtract from the original equation:

$$y_{i,t+h} - \bar{y}_i = \boldsymbol{x}_{\boldsymbol{i,t}}\boldsymbol{\beta} - \bar{\boldsymbol{x}}_{\boldsymbol{i}}\boldsymbol{\beta} + CFE_i - CFE_i + u_{i,t+h} - \bar{u}_i \tag{3.3}$$

The country specific time invariant effect has thus been eliminated. Again, the reasoning behind country-demeaning is the same for eliminating the time specific country invariant effects. While it is not entirely clear whether the authors of the original paper use a dummy variable approach or a demeaning approach, this thesis estimates the second approach using the 'plm' package of the R programming language (Croissant *et al.* 2020). This will allow to report a more reliable R-squared, since the goodness fit of the model will not take explicitly into account the effect of the country and time dummy variables.

Another source of potential bias which has to be dealt with is the dynamic bias. The authors of the paper to be replicated do not treat or mention it extensively, and therefore it might be valuable to investigate into it a bit further here. Note that in equation (3.1) the lags of the dependent variable are included among the dependent variables. A key assumption in estimating panel data equations is that of strict exogeneity:

$$\mathbb{E}(u_{i,t}|X) = 0 \tag{3.4}$$

where X is the matrix of independent variables. However, as Wooldridge (2002) notes, when we include lags of the dependent variable, this assumption is necessarily broken. For simplicity, let us have a model: $y_{i,t} = \beta_1 y_{i,t-1} + u_{i,t}$. Then:

$$\mathbb{E}(y_{i,t}u_{i,t}) = \beta_1\mathbb{E}(y_{i,t-1}u_{i,t}) + \mathbb{E}(u_{i,t}^2) \tag{3.5}$$

From the assumption $\mathbb{E}(y_{i,t-1}u_{i,t}) = 0$, we get:

$$\mathbb{E}(y_{i,t}u_{i,t}) = \mathbb{E}(u_{i,t}^2) > 0 \tag{3.6}$$

Therefore the strict exogeneity assumption, which assumes that the independent variables are uncorrelated with the disturbance for all periods, is broken. This means that the estimates of the coefficients are biased, which has to be kept in mind. However, the consistency of the local projection method has been established by its author, Jorda (2005), through Monte Carlo simulations. A paper by Herbst & Johannsen (2020) then shows, both theoretically and through simulations, that the size of the bias decreases with the time dimension, $T \longrightarrow +\infty$, and increases when projections are done further into future, i.e. with large h. Furthermore, the authors claim that the size of the bias is independent of the size of the cross-country dimension. For accurate estimates, it is necessary to obtain as large T as possible, while estimating equations with large h might not be relevant. Moreover, these results assume that the variables are covariance stationary. This is probably one of the reasons why Bordon *et al.* (2016) do not use the dependent variable in the form proposed by Jorda (2005), i.e. simply $y_{i,t+h}$, but instead subtract $y_{i,t-1}$ from it to eliminate a potential unit root. Also, the lags of the dependent variable are in the form of first differences. While it is reasonable to assume stationarity for the data included in the model, it might perhaps be of value to test these assumptions rigorously, especially since the authors of the original paper did not do so. We decided to test the stationarity of the employment first-difference, the output first-difference, the dependent variables and the output gap. The other variables are dummies or have sparse values. To test unit roots in a panel data setting, the standard approaches used for simple time-series have to be modified. There are basically two main approaches, with many different specific tests. The first approach assumes independence in the cross-sectional dimension. This includes the Choi test, which combines the p-values from unit root tests (Augmented Dickey-Fuller tests) from the separate time-series in the following way (Kleiber & Lupi 2011):

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(p_i) \tag{3.7}$$

Under the null hypothesis: $Z \sim^a N(0, 1)$. However, in a macroeconomic setting such as ours, cross-sectional independence can hardly be guaranteed. A random sample cannot be drawn in the classical sense of the word. There are often dependencies between e.g. the employment rate of one country with the employment rate of a neighbouring one. To correct for this, Hartung (1998) proposed a correction in the following form (Kleiber & Lupi 2011), where ρ is the correlation between p_i:

$$Z = \frac{1}{\sqrt{N(1+\rho(N-1))}} \sum_{i=1}^{N} \Phi^{-1}(p_i) \tag{3.8}$$

Again, under the null: $Z \sim^a N(0,1)$. Note that since ρ has to be estimated, the actual calculation of Z is more complex.

We apply the Hartung (1998) test on our data, with the underlying Augmented Dickey Fuller tests applied with a maximum of one lag (because of annual frequency of the data and the limited time dimension). Whether one lag or no lag is applied is determined automatically using the Akaike Information Criterion. Table 3.1 shows the Z-statistics and the p-values of these tests performed for the output gap variable, the lagged first difference in employment and employment as the dependent variable for $h = 0, 1, 2, 3, 4, 5$. Since all of the p-values are very low, the null hypotheses of unit roots can be rejected on the 95% significance level.

	Gap	$\Delta empl.$	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
Z-statistic	-5.08	-4.99	-4.56	-12.41	-8.54	-3.24	-3.35	-2.30
p-value	$1.85 \cdot 10^{-7}$	$3.10 \cdot 10^{-7}$	$2.51 \cdot 10^{-6}$	$1.20 \cdot 10^{-35}$	$6.89 \cdot 10^{-18}$	$6.00 \cdot 10^{-4}$	$3.97 \cdot 10^{-4}$	0.011

Table 3.1: Panel unit root tests: Output gap and employment

For the GDP as the dependent variable on the other hand, it can be seen in table 3.2 that the unit root hypothesis can be rejected for the first five variables at the 95% level. However, for $h = 4$ it is only possible to reject the unit root hypothesis at the 90% level. For $h = 5$, the p-value is even slightly higher than 10%. Therefore by using these variables, we run the risk of including a non-stationary variable in our regression, therefore breaking the consistency of the local projections method. Even though we do include $h = 4$ and $h = 5$ in our estimations, it is necessary to keep this danger in mind. However, since there is a separate regression for each h, the inclusion of these periods will not influence the estimations for $h < 4$ and the estimation of those equations is valid.

	ΔGDP	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
Z-statistic	-6.23	-5.35	-10.64	-4.39	-2.81	-1.45	-1.25
p-value	$2.30 \cdot 10^{-10}$	$4.39 \cdot 10^{-8}$	$1.01 \cdot 10^{-26}$	$5.67 \cdot 10^{-6}$	$2.47 \cdot 10^{-3}$	0.074	0.105

Table 3.2: Panel unit root tests: GDP

Bordon *et al.* (2016) also mention that they use the robust standard errors proposed by Driscoll & Kraay (1998). It is sensible to follow this approach in this case, since the errors are robust not only to heteroskedasticity and time-dimension serial correlation, but also cross-sectional correlation. As we have stated in the previous paragraphs on testing stationarity, since this is a macroeconomic study concentrating on OECD countries, the sample is not entirely random in the cross-country dimension. Driscoll & Kraay (1998) show that failing to account for that can cause serious bias in the standard errors. Thus also t-statistics and p-values would be rendered invalid.

3.2 Data

The main contribution of this paper is the extension of the original dataset in both the country and the time dimension. Furthermore, some of the variables have been calculated differently from the Bordon *et al.* (2016) paper. This section covers and explains those differences, but let us note that we were not able to contact the authors of the original paper to check their data - therefore there might also be some further differences that were not obvious from the paper. To ensure we do have a similar model to that of Bordon *et al.* (2016) and we are indeed testing its robustness, we will try to get closer to the original results in the robustness checks chapter.

First, let us look at how the issue of reform measurement, which has been described extensively in the literature review, was addressed in the paper by Bordon *et al.* (2016). The authors use two OECD indicators, which have already been described in the literature review: the Regulation in Energy, Transport and Communications (ETCR) indicator as a proxy for a product market reform indicator, and the Employment Protection Legislation (EPL) indicator for labour market reforms. The latter is restricted only to the protection regulation relating to regular (not temporary) workers. From these indicators, the authors calculate the reforms shocks in the following way. First, they take the first difference of the indicators:

$$difference_{i,t} = indicator_{i,t} - indicator_{i,t-1} \tag{3.9}$$

for each $i = 1, \dots, N$ and $t = 1, \dots, T$. Following that, they calculate the reform shocks like so:

$$reformshock_{i,t} = \mathbb{1}[difference_{i,t} \leq -SD] \tag{3.10}$$

for each $i = 1, \dots, N$ and $t = 1, \dots, T$, where $\mathbb{1}[.]$ is the indicator function and SD is the standard deviation of all of the data for the difference of the given indicator. For product market reforms, the threshold is two standard deviations instead of one as for the labour market reforms. The authors justify this by the fact that there are far fewer labour reforms than product market reforms. While this is true and we follow this approach in this part of the thesis, the setting of this threshold is very arbitrary, as we have mentioned in the literature review. The thesis will address this issue in the robustness analysis, when we instead use data from a narrative database. The reform data is available for the period 1985-2013 for the labour market reforms and 1980-2013 for the product market reforms.

Unemployment data is taken from the World Economic Outlook database (IMF 2019), which collects data on unemployment mostly from national statistical offices. It is expressed as a percentage of the total labour force of the given country. Then, employment is calculated as: $E = 100 - U$.

To get a more complete picture, it seems reasonable to use not only the employment, but also the GDP as the dependent variable. The source of our data is the Quarterly National Accounts OECD database (OECD 2014). It contains quarterly, seasonally adjusted data on the GDP, which is in volume estimates with base year 2015 and expressed in millions of US dollars at annual levels. It is computed using the expenditure approach. Data with quarterly frequency is used so as to use the same estimates of GDP as when calculating the output gap (see section 5.2 of the robustness checks). For the purposes of using it as the dependent variable, annual means are taken and the logarithm of the resulting vector is calculated. From this value, the dependent variable values are calculated for each h in the same way as for the employment. When estimating the effect on the GDP, we also replace the lagged first differences in employment with the lagged first differences in the logarithm of the GDP.

Furthermore, Bordon *et al.* (2016) leave certain countries out of the sample, namely countries with a population smaller than five million. In regressions concerning the product market reforms, the authors also exclude countries where no large reforms occurred. For the sake of accuracy, these restrictions are also applied in the next part of the thesis, when we aim to get as close as possible to the results of Bordon *et al.* (2016). However, there is no clear reason why the dataset should not contain countries with a population lower than five million. If there are some particularities of those countries that might influence the results because of the size of the country, they should be more or less time invariant and should be captured by the country fixed effects. Therefore we do include those countries, as opposed to the original study. Any further restrictions on the dataset have also been lifted, except for the requirement of the countries being members of the OECD and having data for the OECD indicators. Moreover, even though the authors claim to have used a sample of OECD countries, they also included Russia, Indonesia and South Africa, which are not members of this organisation. These have not been included in this replication. The structure of these economies is quite different to those of the OECD countries and Bordon *et al.* (2016) do not provide justification for including them. Table 3.3 shows the difference between the samples in the cross-country dimension.

Table 3.3: Sample comparison

New estimation sample			
Australia	Austria	Belgium	Canada
Chile	Czech Republic	Denmark	Estonia
Finland	France	Germany	Greece
Hungary	Iceland	Ireland	Israel
Italy	Japan	Korea	Latvia*
Lithuania*	Luxembourg	Mexico	Netherlands
New Zealand	Norway	Poland	Portugal
Slovak Republic	Slovenia	Spain	Sweden
Switzerland	Turkey	United Kingdom	United States*
Bordon *et al.* (2016) sample			
Australia	Austria	Belgium	Canada
Chile	Czech Republic	Denmark	Finland
France	Germany	Hungary	Indonesia**
Israel	Italy	Japan	Korea
Mexico	Netherlands	Poland	Portugal
Russia**	Slovak Republic	South Africa	Spain
Sweden	Switzerland	Turkey	United Kingdom
United States**			

*The OECD database does not provide data on Latvian, Lithuanian and US product market reforms.
**Countries that were excluded in the product market reform regressions by Bordon *et al.* (2016).

For the time dimension, data is added for the dependent variable up to 2018. Unfortunately, data for the indicators is available only up to 2013. Nevertheless, extended data for the dependent variable is also very useful, since for $h > 0$, future data on the dependent variable is needed. Thus with this data extension, for any h, the data has the same size. For $h > 0$, more periods are added as opposed to the original study. As noted in the previous chapter, this might reduce the dynamic bias. Furthermore, the effects of reforms that were undertaken in the aftermath of the global financial crisis are more extensively captured by this modification.

The data for the output gap in Bordon *et al.* (2016) was taken from the World Economic Outlook database (IMF 2019). However, data for several OECD countries are lacking in the database. The authors of the original paper seem to have included certain of these countries in their sample anyway and it is not very clear where the data was taken from. To remedy this issue, we estimate the output gap ourselves using the method proposed by Hamilton (2018). This method and its comparison to the WEO database and the Hodrick-Prescott filter is further discussed in the robustness checks chapter (section 5.2), where we look at the influence of output gap measurement on the results.

Data on banking crises used by Bordon *et al.* (2016) is taken from the dataset compiled by Laeven & Valencia (2013). This thesis uses an update (Laeven & Valencia 2018). The two variables that the paper used from this dataset are the banking crisis dummies and the output loss during crisis data. It is unclear how exactly the authors used this data. It seems reasonable to include the banking crisis dummies for each year of the duration of the banking crisis. However, the dataset includes a separate table with only the starting years of crises and it is an option to only include those. We decided to use the former approach, however the authors might have used the latter.

Laeven & Valencia (2013) compute the output loss as the cumulative output gap (using the Hodrick-Prescott filter (Hodrick & Prescott 1997)) for three years after a crisis. However, since for each crisis this means only one data point, the resulting data is very sparse and it is unclear how this affects the estimation. Instead, this thesis uses a different approach: for each year from the start of the crisis to five years after its end, the cumulative sum of the output gap is calculated using the Hamilton (2018) filter.

Chapter 4

Results

This chapter shows the new results as compared to the original paper written by Bordon *et al.* (2016). Tables with coefficients, standard errors, t-statistics, p-values, numbers of observations and adjusted R-squareds are presented, as well as graphs of the impulse-response functions with approximate 95% confidence intervals. The authors of the original study provide only the estimates, the t-statistics and the number of observations. The standard errors and the p-values have thus been calculated from the provided values. The p-value has been calculated with the assumption that nine parameters had to be estimated, thus the degrees of freedom are presumed to be equal to $n.observations - 9$. Nevertheless, the sample size is quite large, so the potential imprecision is low. Accordingly, the confidence intervals have been calculated using the standard normal distribution quantiles. For our new estimates, adjusted R-squared has also been calculated. Note that this measure is largely influenced by the fact that the output gap is included as one of the controls, and it was already shown by Okun that there is a strong relationship between unemployment and the output gap (Okun 1963). Indeed, the output gap has been both practically and statistically significant in most of the replication regressions.

Table 4.1 shows the results of the regressions performed with the labour market reform shocks as the reform variable and compares them to the results of Bordon *et al.* (2016).

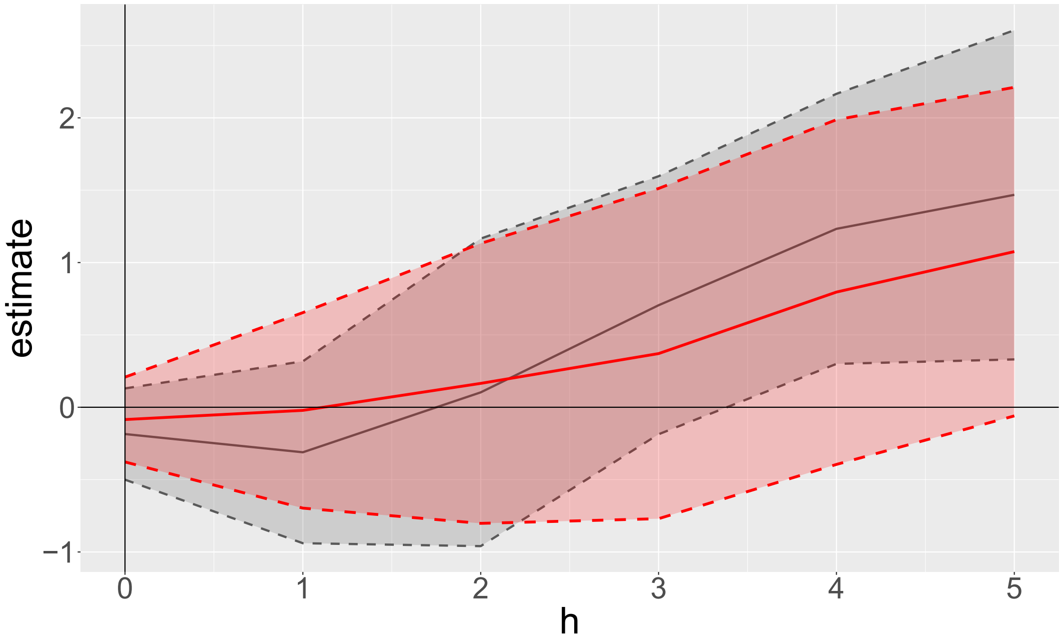
Table 4.1: Effects on employment: Labour market reforms, results comparison

h	Estimate	Standard error	t-statistic	p-value	N. of observations	Adjusted R^2
			New results			
0	-0.085	0.150	-0.567	0.571	745	0.426
1	-0.021	0.345	-0.062	0.950	745	0.337
2	0.165	0.494	0.334	0.738	745	0.219
3	0.371	0.582	0.637	0.524	745	0.139
4	0.796	0.608	1.310	0.191	745	0.106
5	1.076	0.579	1.857	0.064	744	0.098
			Bordon *et al.* (2016)			
0	-0.185	0.161	-1.150	1.749	555	-
1	-0.311	0.321	-0.970	1.668	555	-
2	0.103	0.542	0.190	0.849	555	-
3	0.705	0.455	1.550	0.122	526	-
4	1.233	0.476	2.590	0.010	497	-
5	1.468	0.580	2.530	0.012	468	-

Table shows the results of the replication with extended time and country dimension with respect to the results of Bordon *et al.* (2016). The estimates show the effects of labour market reforms on employment at h periods in the future.

Except for $h = 5$ at the 90% significance level, none of our results are significant, as opposed to the results of Bordon *et al.* (2016), which show significance even for $h = 4$ even at the 95% significance level. In terms of signs of the results, both results indicate short-term costs and long-term benefits, with the effects being negative for the zeroth and first periods after the reform occured and positive afterwards. However, the magnitude of those effects is much larger, except for the second period, in the case of the original study, with the difference being as large as around 40% for the last period. Visually, this is confirmed by figure 4.1, which shows the graphs of the impulse-response functions for both the original results (in grey) and our results (in red), with 95% confidence intervals. While the difference is practically significant, the new results are still within the original confidence intervals.

Figure 4.1: Effects on employment: Labour market reforms, results comparison



An impulse-response function comparison. The red curve represents the new estimates, while the Bordon *et al.* (2016) estimates are shown in grey. The hued regions represent the respective 95% confidence intervals.

When looking at table 4.1 and figure 4.1, it must however be concluded that the evidence for the hypothesis that structural reforms bring significant short-term costs and long-run benefits is quite weak. No significant results have been captured by our replication at the standard 95% level. It is possible that significant results could appear with larger h, however, as we have noted in the methodology section, the local projections method can be unreliable for large values of h.

Now for the product market reforms. Similarly to table 4.1, table 4.2 shows the results of the regressions performed with the product market reform shocks as the reform variable and compares them to the results of Bordon *et al.* (2016).

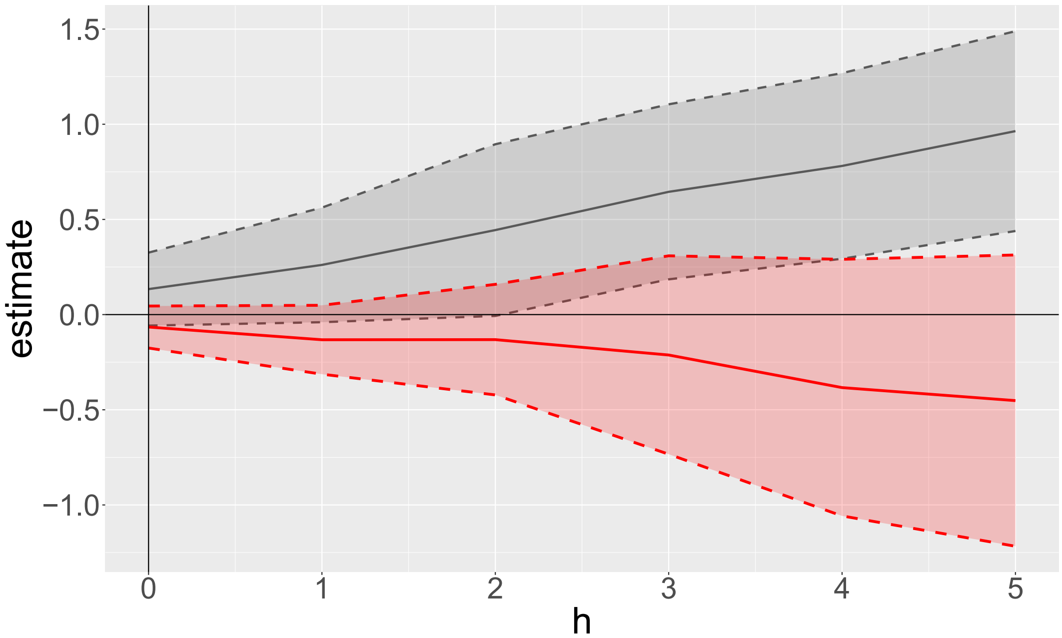
The new results are fundamentally different from those of Bordon *et al.* (2016), with the difference being even stronger than in the case of the labour market reforms. In the new results, no estimate is statistically significant at any reasonable level. Furthermore, the signs for all periods are negative for the new results, while the authors of the original study report only positive effects for all periods. Figure 4.2 shows that all of the estimates fall outside the Bordon *et al.* (2016) confidence intervals, as opposed to the labour market reforms. Moreover, not only are the estimates negative in our results, but the effects are actually getting only more negative over time. Although their magnitude is not large, this trend is in stark opposition to the original results, where the effects are actually strictly increasing over time.

Table 4.2: Effects on employment: Product market reforms, results comparison

h	Estimate	Standard error	t-statistic	p-value	N. of observations	Adjusted R^2
			New results			
0	-0.065	0.056	-1.160	0.246	866	0.382
1	-0.132	0.092	-1.431	0.153	866	0.305
2	-0.131	0.148	-0.887	0.375	866	0.198
3	-0.212	0.266	-0.798	0.425	866	0.128
4	-0.384	0.344	-1.115	0.265	866	0.107
5	-0.452	0.390	-1.158	0.247	866	0.102
			Bordon *et al.* (2016)			
0	0.134	0.098	1.370	0.171	709	-
1	0.261	0.154	1.700	0.090	709	-
2	0.444	0.230	1.930	0.054	709	-
3	0.645	0.235	2.750	0.006	683	-
4	0.781	0.249	3.140	0.002	657	-
5	0.964	0.268	3.600	0.0003	631	-

Table shows the results of the replication with extended time and country dimension with respect to the results of Bordon *et al.* (2016). The estimates show the effects of product market reforms on employment at h periods in the future.

Figure 4.2: Effects on employment: Product market reforms, results comparison



An impulse-response function comparison. The red curve represents the new estimates, while the Bordon *et al.* (2016) estimates and confidence intervals are shown in grey. The hued regions represent the respective 95% confidence intervals.

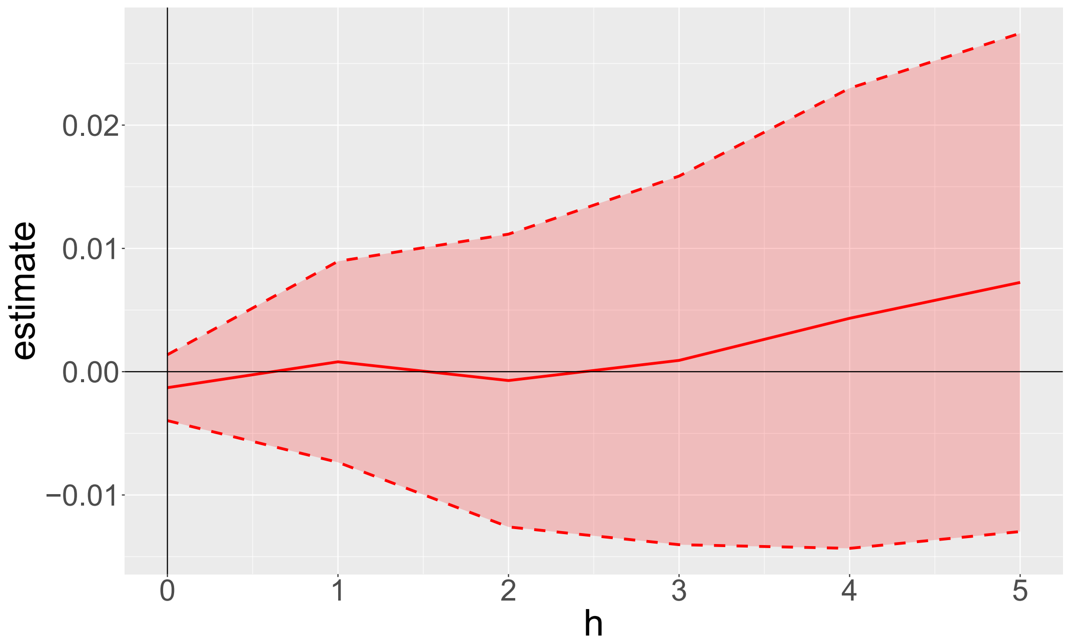
Let us now look at the results with the GDP as the dependent variable. Since it is logarithmically transformed, the exponentiated coefficients are also provided (third column in the tables). These can be interpreted as the multiplicative effect of a reform on the level of GDP before the reform happened. For such small changes as ours, these are almost identical as $estimate + 1$, however it seems clearer to readily provide the exponentiated coefficients. These results are obviously not compared to the results of Bordon *et al.* (2016), since the authors did not calculate the effects on the GDP at all. Table 4.3 shows the effects of labour market reforms on the logarithm of the output, while figure 4.3 shows the impulse-response function with confidence intervals.

Table 4.3: Effects on GDP: Labour market reforms

h	Estimate	$e^{Estimate}$	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	-0.001	0.999	0.001	-0.951	0.342	748	0.878
1	0.001	1.001	0.004	0.193	0.847	748	0.563
2	-0.001	0.999	0.006	-0.118	0.906	748	0.364
3	0.001	1.001	0.008	0.120	0.904	748	0.249
4	0.004	1.004	0.010	0.455	0.649	748	0.184
5	0.007	1.007	0.010	0.702	0.483	747	0.176

The estimates show the effects of labour market reforms on the logarithm of GDP at h periods in the future.

Figure 4.3: Effects on GDP: Labour market reforms



The impulse response-function of the new estimates with the 95% confidence intervals.

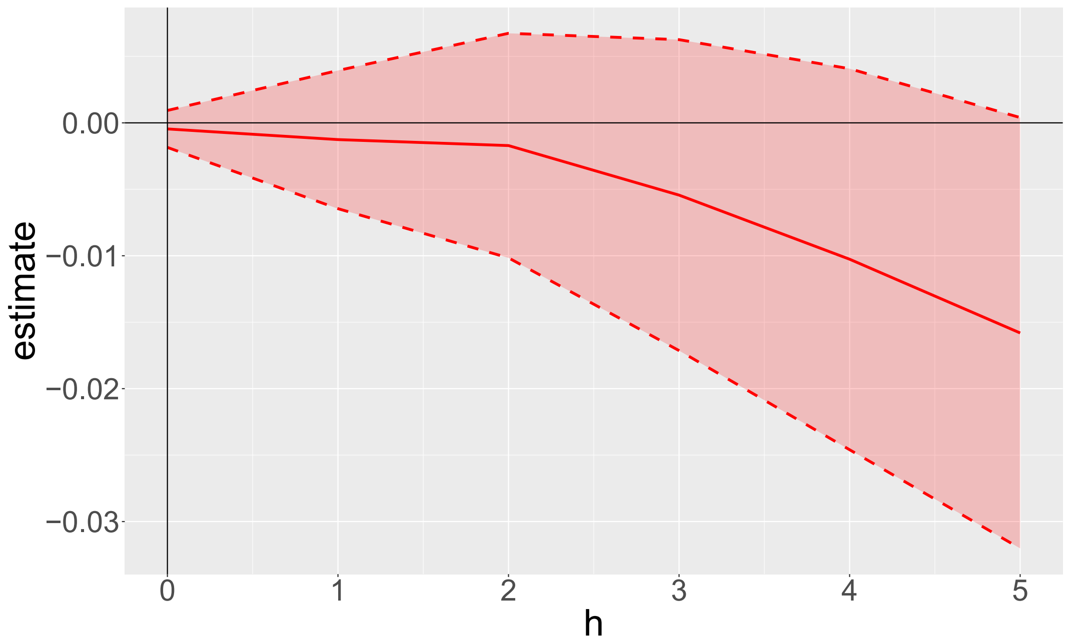
The results are consistent with what was shown for employment. None of the results are statistically significant at any reasonable level. For periods zero to three after the reform, the effect size is very close to zero, then it rises to a 0.7% increase in the GDP in the fifth year after the reform. As compared to the effect on employment, there seems to be no practically significant costs in the earlier periods. The positive effects however show after a longer period. Table 4.4 and figure 4.4 show the effects of product market reforms on the logarithm of the GDP.

Table 4.4: Effects on GDP: Product market reforms

h	Estimate	$e^{Estimate}$	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	-0.0005	0.9995	0.001	-0.644	0.520	946	0.891
1	-0.001	0.999	0.003	-0.475	0.635	946	0.571
2	-0.002	0.998	0.004	-0.397	0.692	946	0.389
3	-0.005	0.995	0.006	-0.911	0.363	946	0.310
4	-0.010	0.990	0.007	-1.403	0.161	946	0.274
5	-0.016	0.984	0.008	-1.912	0.056	946	0.275

The estimates show the effects of product market reforms on the logarithm of GDP at h periods in the future.

Figure 4.4: Effects on GDP: Product market reforms



The impulse response-function of the new estimates with the 95% confidence intervals.

Again, the results are in line with what was estimated for the employment as the dependent variable. The impulse-response function seems to be steadily decreasing and even concave: that is, the effects of the product market reforms are not only negative and decreasing, but are also decreasing faster with higher h. However, the only statistically significant result is the one in the fifth period after the reform, and only at the 90% significance level. As we have said when performing unit root testing, the results for the fourth and the fifth period have to be taken with a grain of salt, since the unit root hypothesis could not be rejected with sufficient certainty. Nevertheless, this points to very different results to the ones established by Bordon *et al.* (2016), who predict positive, increasing and statistically significant effects of product market reforms on the economy.

To conclude this chapter, the replication with the extended dataset shows important changes in the results with respect to the original study. The new calculations cast a shadow of doubt on the often accepted notion that reforms produce significant positive results in a five-year window from the reform. Indeed, no evidence has been found to support this in the new results. Moreover, smaller effect sizes have been calculated than those of Bordon *et al.* (2016). Furthermore, we have shown limited evidence that suggests that product reforms might actually have negative effects on the economy even in the longer term. Three hypotheses arise as to what might have caused the fairly large disparity between our results and those of the original paper:

- Inclusion of more (although still only advanced) economies in the estimation could have changed the results. Choosing the sample selectively and in a non-random way without proper justification could have distorted the original results.

- A different approach to output gap and output loss measurement might have also influenced the results.

- The last possibility is that including a longer period in the estimation might have had the decisive effect. Possibly, effects of the financial crisis on the efficiency of reforms might be at play.

Chapter 5

Robustness checks

In this chapter, we will further test the robustness of the model. First, to check the closeness of the used specification to that of Bordon *et al.* (2016), we restrict the sample. Then, we check how the results react to a different output gap estimation method. Finally, to deal with the arbitrariness of the reform measurement method, we test the robustness of the results with respect to exchanging the reform data for a narrative database of reforms.

5.1 Getting close to Bordon *et al.* (2016)

To confirm whether the replication indeed uses a similar specification as Bordon *et al.* (2016), we will attempt to get closer to the original study. This means primarily restricting our dataset and using the same data for the output gap and the output loss during crisis as the original authors used. Note however that the we were not able to get the exact same dataset as the authors.

In order to replicate the results of Bordon *et al.* (2016), the data has been restricted to the period before 2013, since the authors explicitly state that their dataset ends in 2013. While the reform data is available only up to that year anyway, it is important to restrict the dependent variable in this way, since the regressions are performed for up to 5 leads of the dependent variable into the future (see equation (3.1)).

Further, to get closer in terms of the variables used, the output loss during crisis which we have calculated ourselves has been replaced by the data from Laeven & Valencia (2018). The output gap data has been replaced by the WEO data. However, as we have stated in the previous sections, this data is publicly available only for certain countries from the Bordon *et al.* (2016) sample. This means that the number of countries we include will be actually lower than in the original study this time. Furthermore, we apply the same restrictions on the sample as the original study, most importantly excluding countries with a population smaller than five million and countries without significant reforms. A clear country composition of our sample can be seen in table 5.1.

Table 5.1: Restricted sample

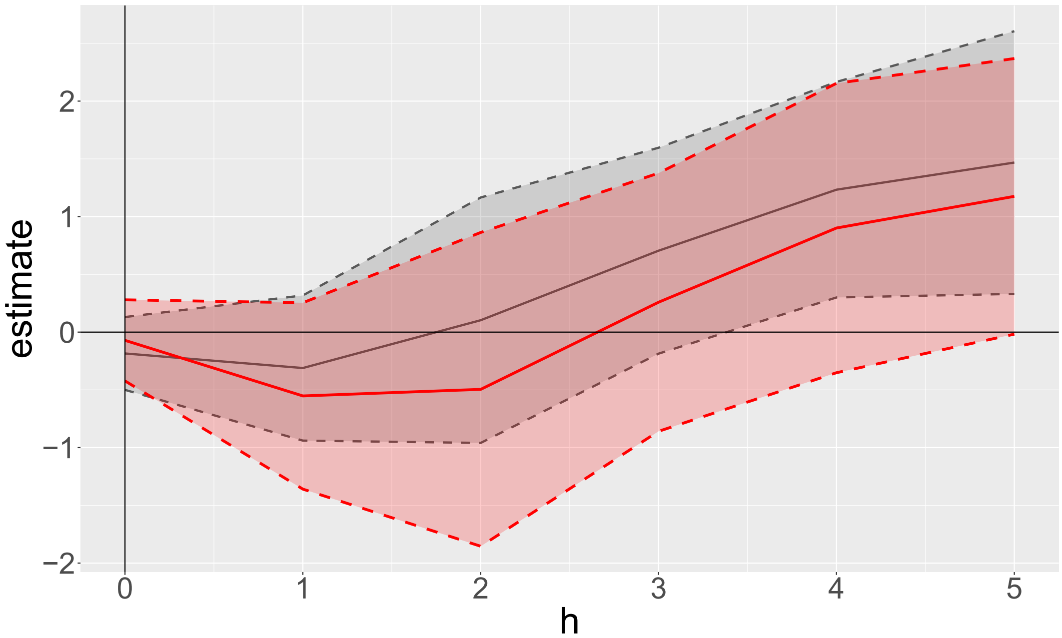
Australia	Austria	Belgium	Canada
Denmark	Finland	France	Germany
Italy	Japan	Korea	Netherlands
Portugal	Slovak Republic	Spain	Sweden
United States*	United Kingdom		

*Exluded from the product market reform regression.

5.1.1 Results

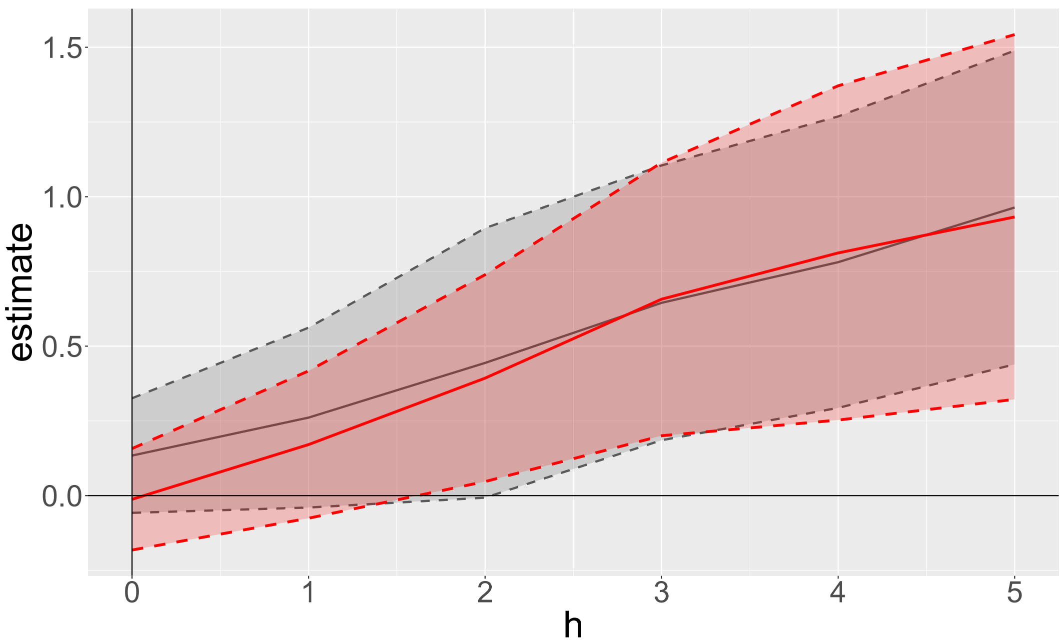
Figures 5.1 and 5.2 show the impulse-response function of the results from this section (in red) in comparison with the original estimates (in grey) with 95% confidence intervals. The graph of the labour market reforms impulse-response is indeed very similar, at least in shape, to the original estimates. Again, the new impulse-response is well within the original estimates. However, the estimates are still somewhat lower than in the original study.

Figure 5.1: Effects on employment: Labour market reforms, restricted sample



An impulse-response function comparison. The red curve represents the new estimates, while the Bordon *et al.* (2016) estimates are shown in grey. The hued regions represent the respective 95% confidence intervals.

Figure 5.2: Effects on employment: Product market reforms, restricted sample



An impulse-response function comparison. The red curve represents the new estimates, while the Bordon *et al.* (2016) estimates are shown in grey. The hued regions represent the respective 95% confidence intervals.

For the product market reforms, the change from the preceding chapter is dramatic. Indeed, this time the impulse-response is well within the confidence intervals of the original study. What is more, the estimates are basically identical for periods 3 to 5. The only real disparity is for the first three periods, with the effects being lower for the new estimates.

To conclude this section, even though the data and the specification is still not identical to Bordon *et al.* (2016), it has been possible to get close enough to the original estimates by restricting the dataset and performing further adjustments. This is especially important in the case of product market reforms regressions which showed fundamental changes in the estimates when the dataset was extended in the preceding chapter. This allows us to say with a degree of confidence that the difference is indeed due to the changes specified in the preceding chapter and not due to some error in our estimation vis-a-vis that of Bordon *et al.* (2016).

5.2 Output gap specifications

The output gap is standardly defined as the difference between the actual output and the potential output. As compared to some other macroeconomic indicators, the role of methodology when estimating the output gap is crucial. This is because potential output cannot be directly measured by statistical offices and has to be instead estimated. Bordon *et al.* (2016) use the output gap variable from the World Economic Outlook database. This data has been compiled by IMF staff using an approach based on the production function (De Masi *et al.* 1997). While the approach is theoretically sound, there is significant variability in the methodology across countries. Furthermore, the data is not available for a large portion of countries, including certain OECD countries. For these reasons, we have instead used the method proposed by Hamilton (2018) to estimate the output gap ourselves. However, there is at least one other statistical approach to estimating the output gap, which has been traditionally used - the Hodrick-Prescott filter (Hodrick & Prescott 1997). Since the output gap seems to be a very important control variable, it could be worth checking the robustness of the model with respect to this estimation method of the output gap as well. In this section, the thesis first compares the two very different, even conflicting methodologies of the Hodrick-Prescott filter and the Hamilton's approach. Then, we compare these methods with the WEO output gap in terms of descriptive statistics and visualisations. Finally, the results of the regressions with the Hodrick-Prescott filter are reported in order to compare them with the preceding results which used the Hamilton's approach and the WEO output gap.

5.2.1 Output gap estimation methods

Firstly, let us look at the Hodrick-Prescott filter. Hodrick & Prescott (1997) define the method in the following way. The main idea is to decompose a GDP time series (or any macroeconomic time series) into its growth and cyclical component:

$$y_t = g_t + c_t \tag{5.1}$$

where y_t is the GDP, g_t is the potential, c_t is the business cycle and $t = 1, \ldots, T$, where T is the size of the time dimension. The size of the components is estimated by solving the following optimization problem:

$$\min_{g_{-1},\ldots,g_T} \left(\sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right) \tag{5.2}$$

where $\lambda \in (0, +\infty)$ is a parameter chosen by the researcher. Hodrick & Prescott (1997) use a value of $\lambda = 1600$ for their quarterly data.

Even though the Hodrick-Prescott filter is widely used, it has several drawbacks, which has been pointed out in a paper by Hamilton (2018) aptly named: "Why you should never use the Hodrick-Prescott filter". The author brings attention to several problems:

1. The HP filter decomposes an econometric series without having regard for the underlying process. Rather, the values produced by it are based on the application of the filter and the choice of λ.
2. Future values of the series influence how the decomposed parts of the time series are estimated. This means that the output gap will be estimated differently for a given year dependent on which future values of the time series the researcher chooses to include.
3. The fact that the choice of the λ hyperparameter is arbitrary.

Instead, the author proposes to use a different estimation method. It consists of estimating the following equation using OLS:

$$y_{t+h} = \alpha + \sum_{i=0}^{p-1} \beta_i y_{t-i} + u_{t+h} \tag{5.3}$$

where $h, p \in \mathbb{N}$ and $t = 0, \ldots, T$. The residuals $\hat{u}_{t+h}$ then provide an estimate of the cyclical component. With a degree of simplification, we can describe the method as follows: the regressors, i.e. the lagged values of the regressand, control for the stable component of the series (growth). The part of the variance which cannot be explained by these is then considered to be the cyclical component.

The author proposes $h = 8$ and $p = 4$ for quarterly data. The choice of h is influenced by the fact that the interest is in short-term cyclical changes and two years (eight quarters) seems to be a reasonable time horizon. Hamilton (2018) also rigorously proves that the method handles non-stationary series integrated of degree d when $d \leq p$. Thus under the proposed values, this method allows for a $I(4)$ time series. However, as the author notes, as opposed to the Hodrick-Prescott filter, the method does not strictly assume this degree of integration, which can thus also be lower than four.

For the purposes of our estimation, we follow the recommendations of both authors. That is, we use $\lambda = 1600$ for the Hodrick-Prescott filter and $h = 8, p = 4$ for the Hamilton's method. Even though the ultimate goal of this thesis is to use the output gap data in an annual panel (since reform data is not available on a quarterly basis), it is preferable to use quarterly time series given that it is desirable to have the best approximation possible. Furthermore, quarterly GDP data is available for a wide array of countries and periods. As in the data for dependent variables, we use the OECD VPVOBARSA GDP time series (OECD (2014), for details on this data, see the above description in the section on data), which we transform using the natural logarithm. Then the estimation proceeds as follows. For each country, we estimate separate models for both methods. Following that, the cyclical component is extracted. Since data on an annual frequency is needed, we calculate annual means of the resulting series for each country.

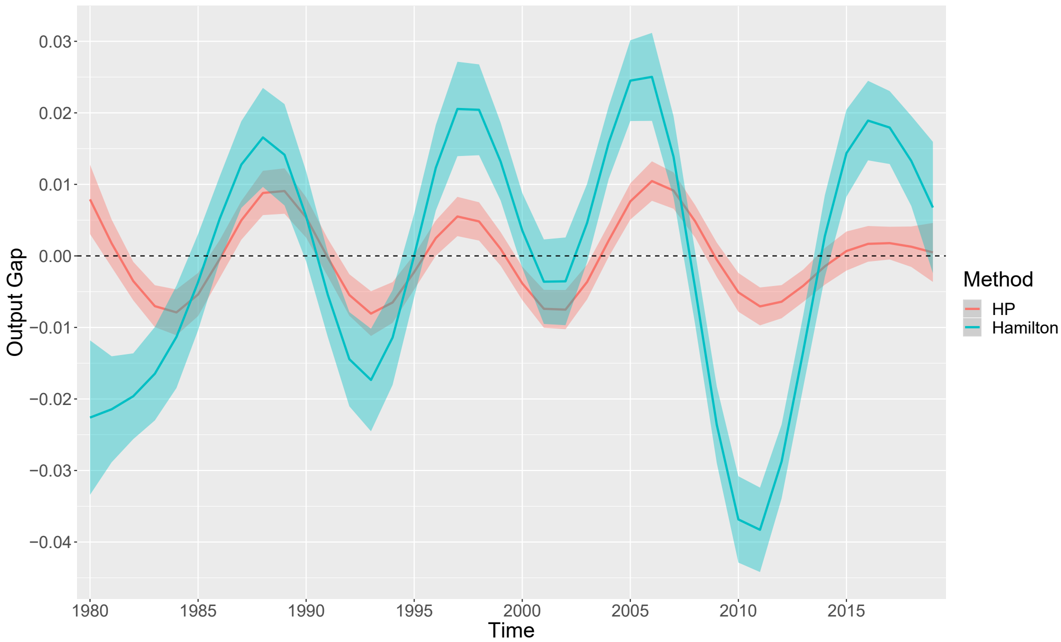
Let us present some summary statistics and a visualisation so that we understand the difference between the results of the two methods in action.

Table 5.2: Output Gap: summary statistics

	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Observations	Correlation
HP	-0.09295	-0.00828	-0.00056	-0.00002	0.00775	0.10878	1291	0.63370
Hamilton	-0.22360	-0.01690	0.00484	0.00006	0.02205	0.21314	1261	0.63370

Table 5.2 shows descriptive statistics of the set of the countries' time series of both methods. When looking at the distance between the mean and the

Figure 5.3: Output Gap: GAM graph summarizing the sample



A Generalized Additive Model graph summarizing the data for the Hamilton and Hodrick-Prescott filters over the whole panel dataset with respect to time, with 95% confidence intervals.

median of the methods, we can conclude that they are all very close to zero and more importantly, close to each other. Therefore the distributions of both the Hodrick-Prescott and the Hamilton data seem to be quite symmetrical around zero. This is a desirable property for output gap data, since it indicates that a growth trend has been successfully isolated. Further, quite a large difference can easily be spotted between the interquartile ranges. The range between the first and third quartiles is much larger in the case of the Hamilton's method. This means that this method tends to estimate depressions and booms as more severe than the Hodrick-Prescott filter. In line with this reasoning, the minimum and the maximum are twice as large in absolute value when using the Hamilton filter. Another factor we might consider is the fact that the number of observations is higher for the HP filter. This is because for a few countries, data on quarterly GDP for certain earlier periods is not available (e.g. Estonia and the Czech Republic were created in the 1990s). While the Hodrick-Prescott filter makes efficient use of all of the periods, the Hamilton's method needs $p+h$ past values to estimate the output gap in the current period (in our case that would equal 12, which then translates into three years). This methodological difference of course results in a higher number of estimated values when using the Hodrick-Prescott filter.

The figure 5.3 shows a smoothed graph of the data. Due to the difficulty of clearly summarizing multiple time series at once, using time as the independent variable and the whole panel of country values of both variants of the output gap seems like a viable option. The graph was created using the Generalized Additive Model technique which is readily accessible through statistical software. This technique uses a sum of various smooth functions f_j which are estimated so that they fit the data as precisely as possible (Hastie & Tibshirani 1986), i.e. the model consists in the following equation:

$$\mathbb{E}(Y|X) = \beta_0 + \sum_{j=1}^{K} f_j(x_j) \tag{5.4}$$

where Y is the dependent variable, $X = (x_1, ..., x_K)$ are the independent variables and β_0 is the intercept. This method also allows for the estimation of confidence intervals. Specifically, the bands around the estimated line in the figure are 95% confidence intervals. From the graph, we can confirm that the Hamilton's method estimates the booms and recessions as more profound than the Hodrick-Prescott filter, at least for our data. This difference is actually often statistically significant, since a large portion of the confidence intervals do not overlap. Also, the size of the HP filter cycle seems to be more regular, whereas there is quite a large degree of heterogeneity in the case of the Hamilton's filter.

To conclude this subsection, while the Hamilton's method seems to be more theoretically sound, the Hodrick-Prescott filter bears the advantage of estimating a slightly higher number of values (this might reduce dynamic panel bias). Furthermore, there is a statistically significant difference for certain periods. Therefore it seems legitimate to not only use the method of Hamilton, but also check the robustness of the models using the Hodrick-Prescott filter.

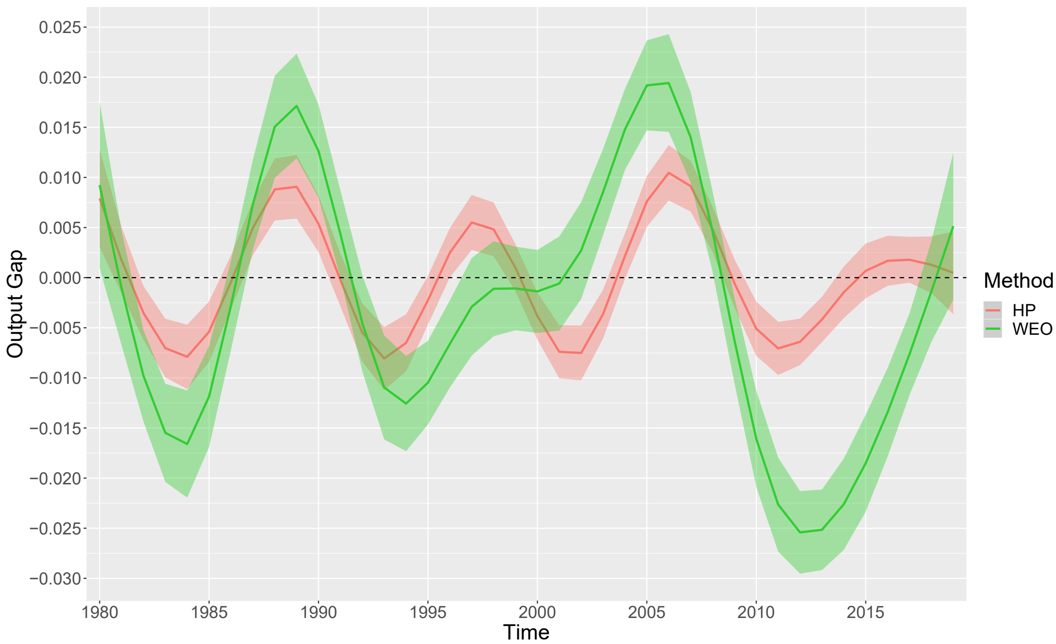
5.2.2 Comparison to WEO output gap

Since Bordon *et al.* (2016) use the WEO output gap in their estimation, it seems legitimate to compare this data with the data obtained from the two methods that have been just described. Figures 5.4 and 5.5 show similar GAM comparisons as in the preceding subsection, this time with respect to the data in the World Economic Outlook database, while table 5.3 shows the descriptive statistics of the output gap from the WEO database, including correlation coefficients with the data from the two methods presented above. From the graphs, it can be noted that the WEO data is in some parts significantly different from both of the methods presented above. With respect to the minimum and maximum values and the IQR, it is situated somewhere between the two methods. We can deduce from the correlation coefficients that the WEO output gap is more aligned with the Hodrick-Prescott data. Furthermore, the number of missing values is drastically higher with the WEO data. This is because the data is entirely missing for certain countries.

Table 5.3: Output Gap - WEO: summary statistics

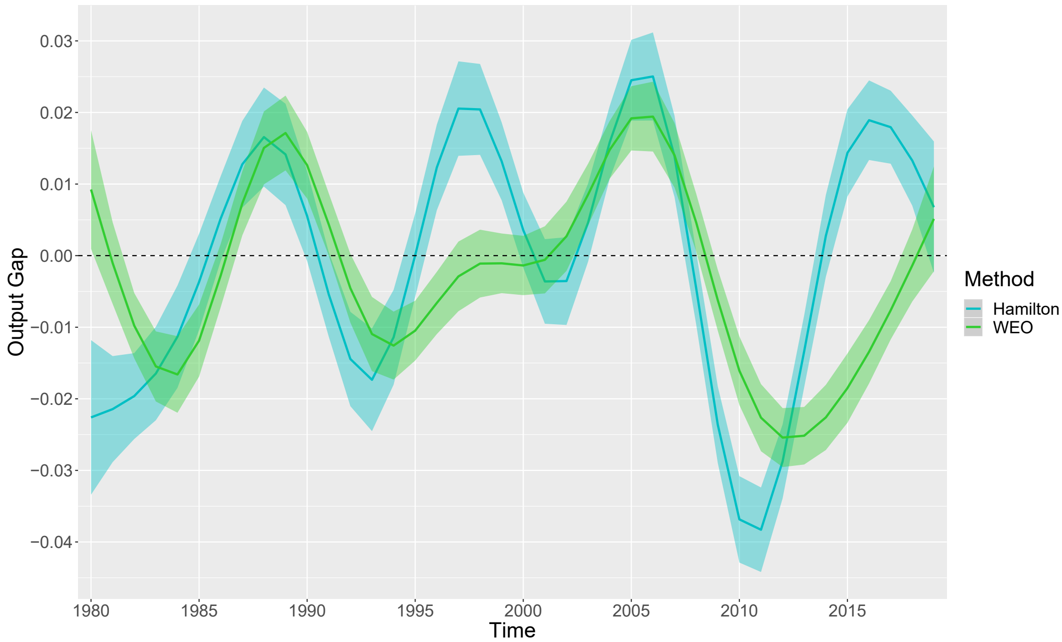
Min	1st Quartile	Median	Mean	3rd Quartile	Max	Obs.	Cor. HP	Cor. Ham.
-0.15810	-0.01626	-0.00326	-0.00298	0.00971	0.11453	935	0.63410	0.48141

Figure 5.4: Output Gap: GAM graph summarizing the sample, HP and WEO



A Generalized Additive Model graph summarizing the data for the Hodrick-Prescott filter and the WEO data over the whole panel dataset with respect to time, with 95% confidence intervals.

Figure 5.5: Output Gap: GAM graph summarizing the sample, Hamilton and WEO



A Generalized Additive Model graph summarizing the data for the Hamilton method and the WEO data over the whole panel dataset with respect to time, with 95% confidence intervals.

5.2.3 Results

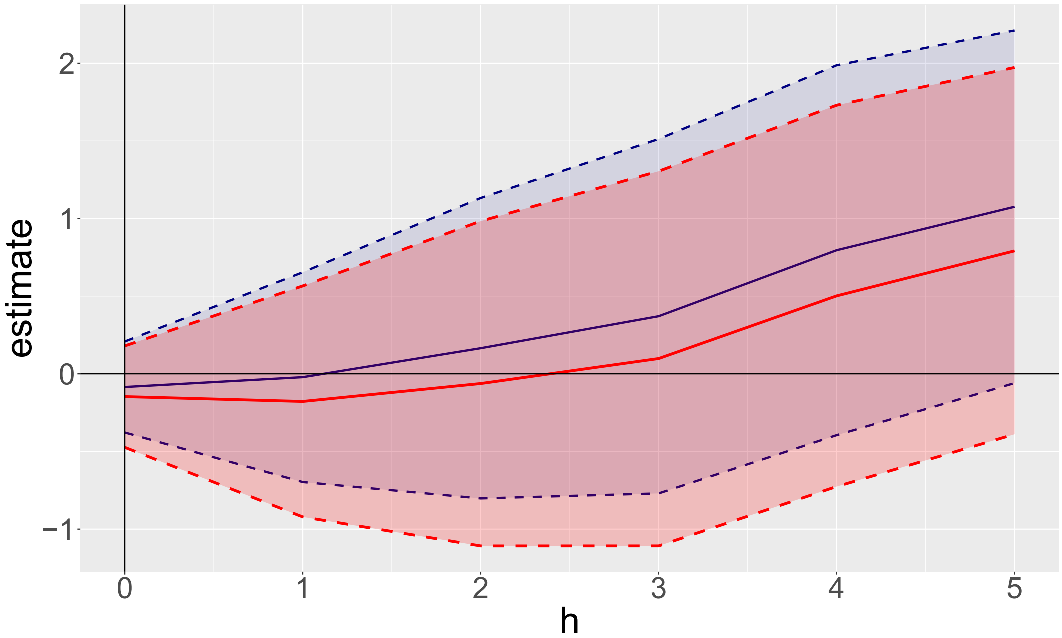
Similarly to the preceding results, table 5.4 shows the results of the regressions performed with labour market reform variable. This time, the Hodrick-Prescott filter has been used to estimate the output gap control variable. Figure 5.6 then shows the graph of the new impulse-response function with 95% confidence intervals (in red) and it compares it with the impulse-response obtained from the main results with the Hamilton's method (in blue).

Table 5.4: Effects on employment: Labour market reforms, HP filter

h	Estimate	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	-0.147	0.167	-0.880	0.379	751	0.410
1	-0.178	0.379	-0.468	0.640	751	0.364
2	-0.063	0.533	-0.117	0.907	751	0.322
3	0.098	0.615	0.160	0.873	751	0.279
4	0.502	0.626	0.801	0.423	751	0.220
5	0.792	0.602	1.314	0.189	750	0.162

The estimates show the effects of labour market reforms on employment at h periods in the future. The Hodrick-Prescott filter has been used to estimate output gap.

Figure 5.6: Effects on employment: Labour market reforms, HP filter



An impulse-response function comparison. The red curve represents the estimates with the HP filter, while the estimates with the Hamilton's method are shown in blue. The hued regions represent the respective 95% confidence intervals.

None of the results are statistically significant - no even in the last period, which was the case when the Hamilton's method was used. Furthermore, the effects are smaller than in the previous case and negative for a longer period of time. However, the shape of the impulse-response function is very similar: it is increasing and also convex for the first four periods. The model therefore still predicts that the effects of reforms are increasing over time. However, to reiterate, there is not sufficient evidence to reject the hypothesis that the effect is zero for any period, as opposed to what Bordon *et al.* (2016) estimate.

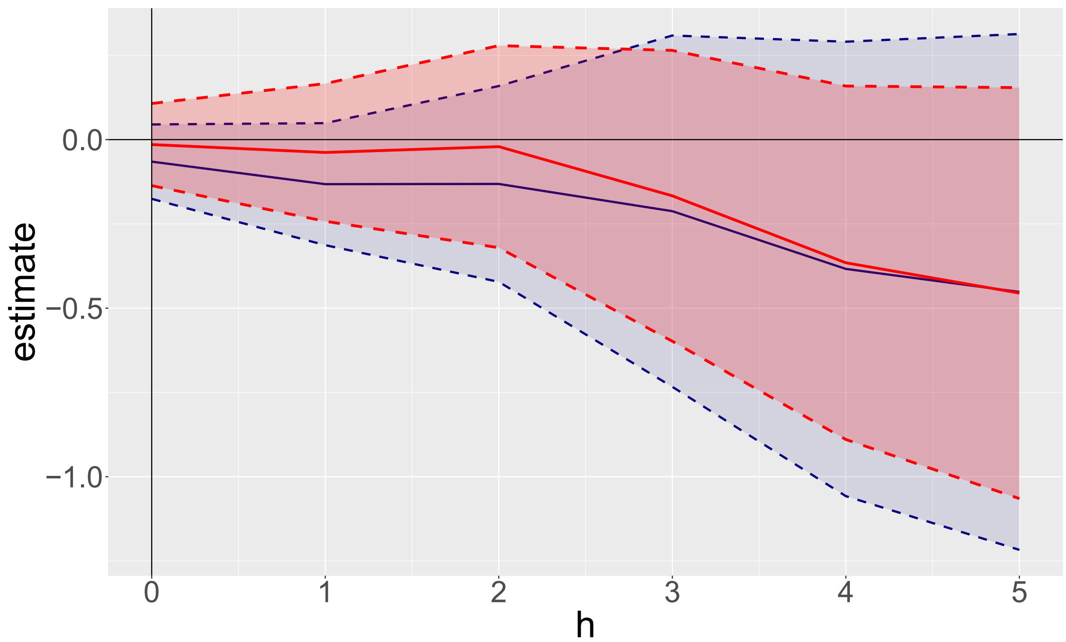
As for the product market reform regressions, table 5.7 shows the estimates and figure 5.7 depicts the graph of the impulse-response function with confidence intervals.

Table 5.5: Effects on employment: Product market reforms, HP filter

h	Estimate	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	-0.015	0.062	-0.236	0.813	882	0.367
1	-0.038	0.104	-0.365	0.715	882	0.342
2	-0.021	0.153	-0.136	0.892	882	0.319
3	-0.167	0.220	-0.757	0.449	882	0.281
4	-0.365	0.267	-1.366	0.172	882	0.224
5	-0.455	0.311	-1.464	0.144	882	0.167

The estimates show the effects of product market reforms on employment at h periods in the future. The Hodrick-Prescott filter has been used to estimate output gap.

Figure 5.7: Effects on employment: Product market reforms, HP filter



An impulse-response function comparison. The red curve represents the estimates with the HP filter, while the estimates with the Hamilton's method are shown in blue. The hued regions represent the respective 95% confidence intervals.

Again, the results are very similar to the ones in the preceding chapter and very different to what Bordon *et al.* (2016) estimated. None of the estimates are statistically significant. As opposed to the results with the Hamilton filter, the effects are very close to zero for the first few periods. For periods four and five, the estimates are practically identical.

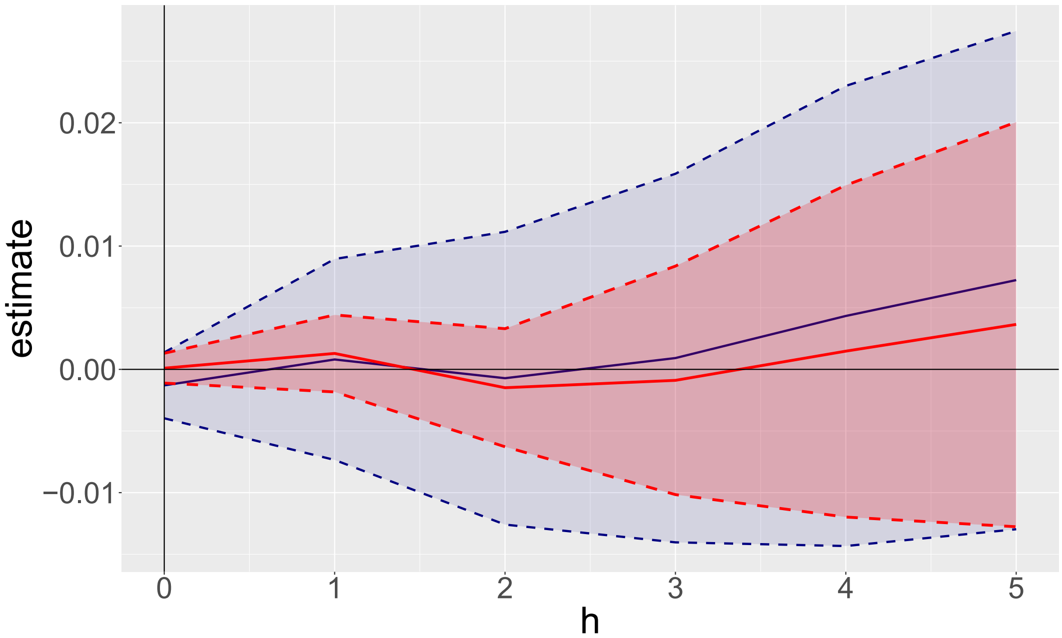
In table 5.6 and in figure 5.8 we can see the effects of the labour market reforms on the logarithm of the GDP, again with the Hodrick-Prescott filter as the estimation method of the output gap.

Table 5.6: Effects on GDP: Labour market reforms, HP filter

h	Estimate	$e^{Estimate}$	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	0.0001	1.0001	0.001	0.155	0.877	748	0.977
1	0.001	1.001	0.002	0.811	0.418	748	0.919
2	-0.001	0.999	0.002	-0.610	0.542	748	0.890
3	-0.001	0.999	0.005	-0.189	0.850	748	0.741
4	0.001	1.001	0.007	0.215	0.830	748	0.556
5	0.004	1.004	0.008	0.435	0.664	747	0.392

The estimates show the effects of labour market reforms on the logarithm of the GDP at h periods in the future. The Hodrick-Prescott filter has been used to estimate output gap.

Figure 5.8: Effects on GDP: Labour market reforms, HP filter



An impulse-response function comparison. The red curve represents the estimates with the HP filter, while the estimates with the Hamilton's method are shown in blue. The hued regions represent the respective 95% confidence intervals.

The pattern is repeating even in this case. The results are not statistically significant and even smaller than in the case when we have used the Hamilton filter. Indeed, the effects seem to be very small and of no actual practical significance.

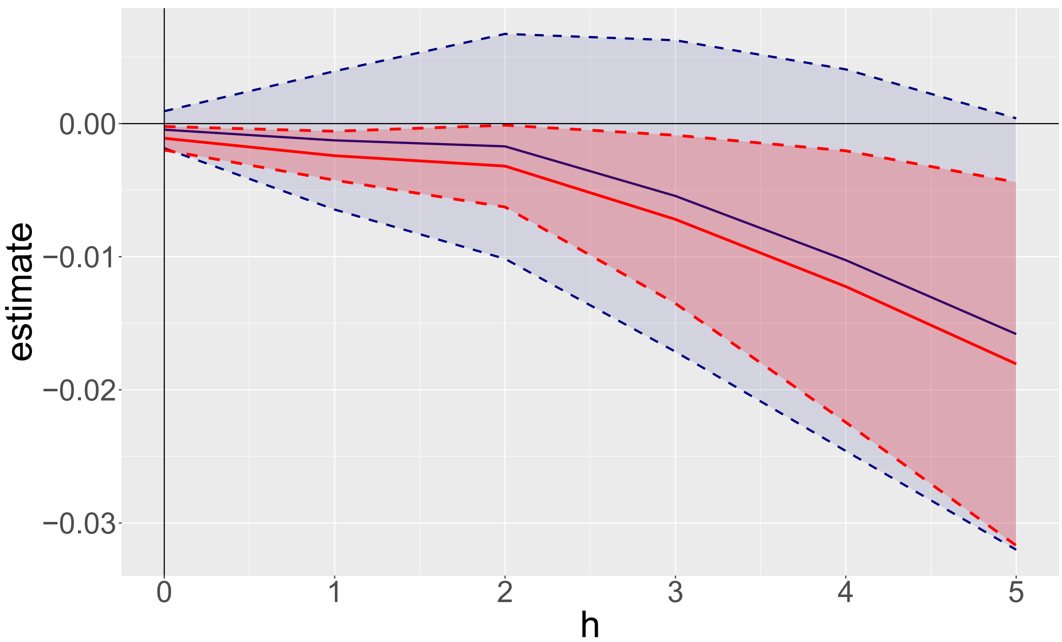
The GDP results for the product market reforms are in table 5.7 and figure 5.9. This time the results differ importantly from the ones before in terms of statistical significance. For all periods, the effect is significantly different from zero at the 95% significance level. Furthermore, the effect is negative for all periods and important in size. For the fifth period after the reform occurs, the GDP falls down by almost 2%. However, it is once more important to remember that for periods four and five, the unit root hypothesis for the GDP dependent variable could not be rejected with sufficient certainty.

Table 5.7: Effects on GDP: Product market reforms, HP filter

h	Estimate	$e^{Estimate}$	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	-0.001	0.999	0.0005	-2.440	0.015	946	0.924
1	-0.002	0.998	0.001	-2.572	0.010	946	0.827
2	-0.003	0.997	0.002	-2.036	0.042	946	0.787
3	-0.007	0.993	0.003	-2.230	0.026	946	0.657
4	-0.012	0.988	0.005	-2.353	0.019	946	0.499
5	-0.018	0.982	0.007	-2.592	0.010	946	0.339

The estimates show the effects of labour market reforms on the logarithm of the GDP at h periods in the future. The Hodrick-Prescott filter has been used to estimate output gap.

Figure 5.9: Effects on GDP: Product market reforms, HP filter



An impulse-response function comparison. The red curve represents the estimates with the HP filter, while the estimates with the Hamilton's method are shown in blue. The hued regions represent the respective 95% confidence intervals.

In conclusion, the Hodrick-Prescott filter has indeed had visible effects on the results of our estimations. Summarizing the results, it can be stated that with the Hodrick-Prescot filter, the results tend to be estimated with lower effect sizes. Most importantly, for product market reforms, the estimation has shown negative and statistically significant effects on the GDP for all of the periods. It must be kept in mind that the Hamilton's method seems to be more theoretically sound and avoids the numerous faults of the Hodrick-Prescott filter. The results from the preceding chapter are therefore probably more reliable. However, the results here show that the estimation method of the output gap has important effects on the results of the final estimates. The output gap seems to be a very important control variable which should be measured carefully. Choosing a wrong method could result in the bias of the final estimates and more attention should be paid to this issue in the relevant literature.

5.3 Narrative dataset

In this section of the robustness checks, we replace the reform variable data by a narrative dataset compiled by Duval *et al.* (2018). As we have already discussed, the detection of shocks based on reform indicators is quite arbitrary, since there is no rule on how to set the relevant threshold. This extension is supposed to remedy this: to record a reform, data has to meet more stringent criteria. Specifically, the authors of the database detect a reform based on the *OECD Economic Survey*, which are publications on the current state of the economy of a given set of countries. Duval *et al.* (2018) record a reform if there has been "strong normative language" indicating it in the *OECD Economic Survey*, if a reform has been mentioned several times across different issues of the *OECD Economic Survey* over time, or if a relevant indicator of regulation has shown a change in "the fifth percentile of the distribution of the cumulative change in the indicator over three years".

Aside from the change in reform measurement, to take a complete picture, data for labour reforms will not include only a measure of Employment Protection Legislation for regular workers, but also of Employment Protection Legislation for temporary workers, unemployment benefits replacement rates and unemployment benefits duration, which are all part of the database. To record if a labour market reform has occurred, a dummy is created to indicate whether any of those categories records a positive reform shock. For product market reforms, the composition is similar, except for the fact that the underlying categories are reforms in the electricity, gas, telecommunications, postal services and rail, air and road transport sectors. This corresponds to the span of the ETCR indicator used in the previous sections.

Furthermore, since the database also contains negative reform shocks (reform reversals), a separate variable will be created to account for this. This is because, as it is stated by Jorda (2005), the local projection method makes the assumption that the effect of a negative shock is completely symmetrical with respect to a positive reform shock. The reason for this is that in case of a negative reform, $\beta_{1,h}$ (equation (3.1)) is multiplied by -1 and in case of a positive one, by 1. Therefore the effect is precisely the opposite. This is too strong of an assumption to make.

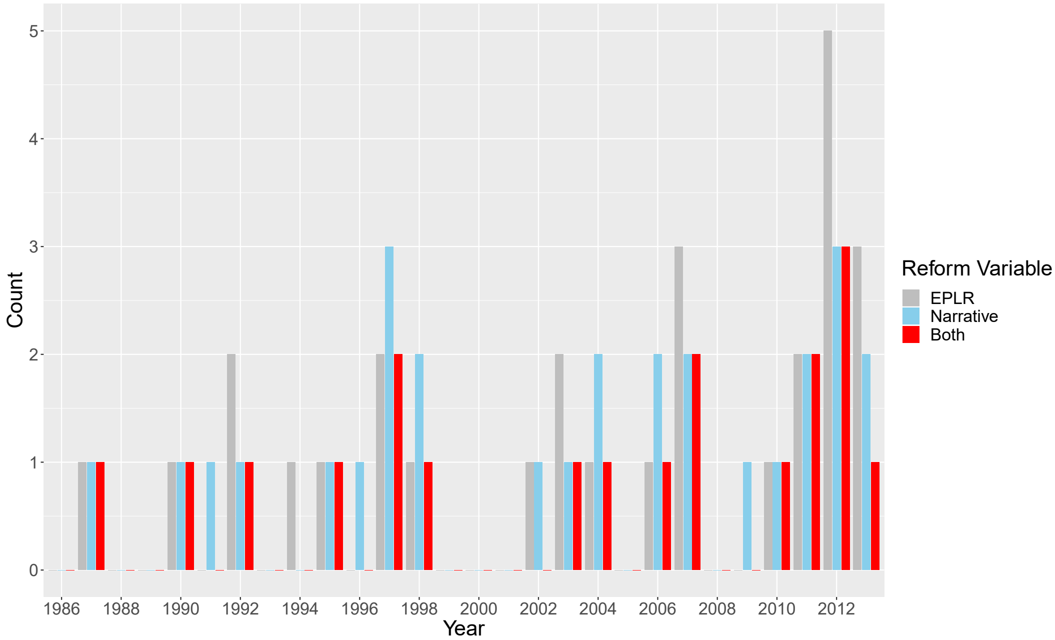
To at least partially visualise the difference between the two approaches of measuring reforms and the the degree to which they are similar, we present figures 5.10 and 5.11. The bar plots show the count of reforms in a given year across the whole panel. The colours represent either the indicator data, the narrative data, or whether a reform happened in both datasets. To actually summarize the differences between the datasets, two restrictions have been put in place:

1. If a given country and year combination is not present in any of the two datasets, the reform is not counted.
2. For the labour market reforms, only Employment Protection legislation for regular workers is used from the narrative dataset.

In the labour market reform plot, it can be seen that the actual number of reforms for both methods is quite low. Whether more reforms are found in one dataset or the other is largely variable, however, for the last few periods, a much larger number is identified by the Bordon *et al.* (2016) approach than by the narrative dataset.

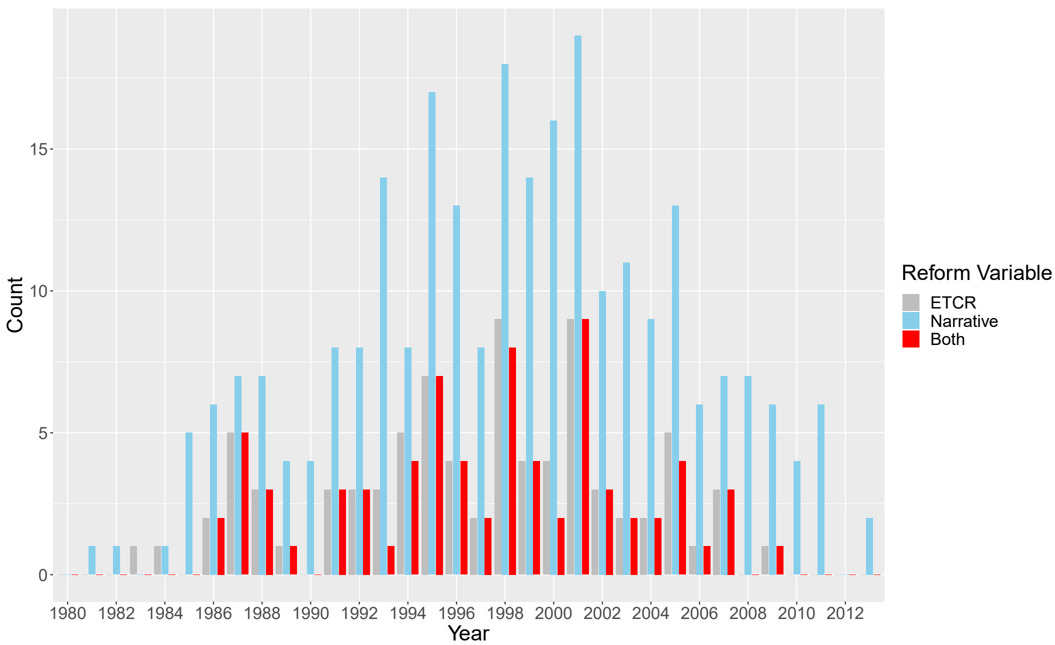
On the other hand, for the generally numerous product market reforms, the larger number of identified reforms is captured by the narrative dataset. The quantity is often as much as two times higher. This might be caused by the fact that in case of product market reforms, Bordon *et al.* (2016) used a higher threshold to identify a reform than in the case of the labour market reforms. Therefore there might be sufficient evidence to indicate a reform in the *OECD Economic Survey*, while the change in the indicator might not be large enough.

Figure 5.10: Labour market reforms: comparison of datasets



A bar plot showing the counts of reforms over time. Grey bars show data based on indicators, blue bars represent narrative data and red bars show their intersection.

Figure 5.11: Product market reforms: comparison of datasets



A bar plot showing the counts of reforms over time. Grey bars show data based on indicators, blue bars represent narrative data and red bars show their intersection.

5.3.1 Results

This is where the results with the narrative dataset are presented. The change in the data is the only change with respect to the main results, note however that this means that the number of observations might be different - the set of missing variables is slightly different in both datasets.

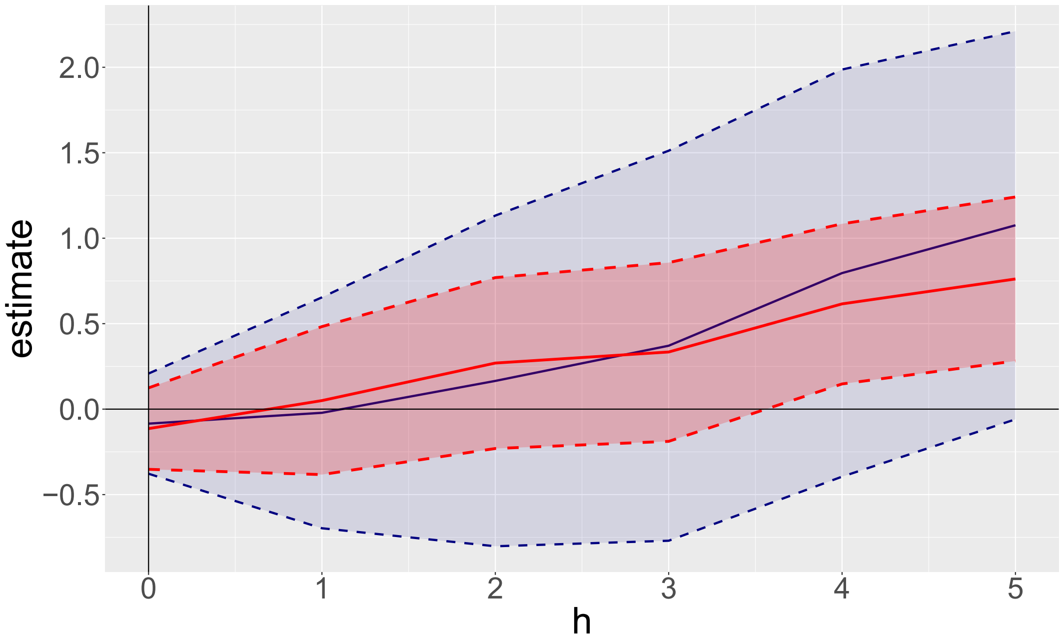
Table 5.8 shows the results of the new estimation and the impulse-response function with the 95% confidence intervals is depicted in figure 5.12. The new estimates are presented in red, while the results from chapter 4 are presented in blue.

Table 5.8: Effects on employment: Labour market reforms, narrative dataset

h	Estimate	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	-0.114	0.122	-0.937	0.349	747	0.408
1	0.050	0.221	0.226	0.821	747	0.353
2	0.270	0.255	1.057	0.291	747	0.243
3	0.334	0.267	1.253	0.211	747	0.160
4	0.616	0.239	2.577	0.010	747	0.121
5	0.762	0.245	3.111	0.002	747	0.098

The estimates show the effects of labour market reforms on employment at h periods in the future, with the narrative dataset as the reform data.

Figure 5.12: Effects on employment: Labour market reforms, narrative dataset



An impulse-response function comparison. The red curve represents the estimates with the narrative dataset, while the estimates with the data based on indicators are shown in blue. The hued regions represent the respective 95% confidence intervals.

The results show a similar evolution of the effects to the results with the non-narrative dataset. However, the standard errors are much lower than before and for years four and five, the new estimates are statistically significant at the 95% level. While the statistical significance is higher, the effect size is lower. When compared to Bordon *et al.* (2016), effect for the fifth year after the reform in the new estimates is around a half of what they estimated.

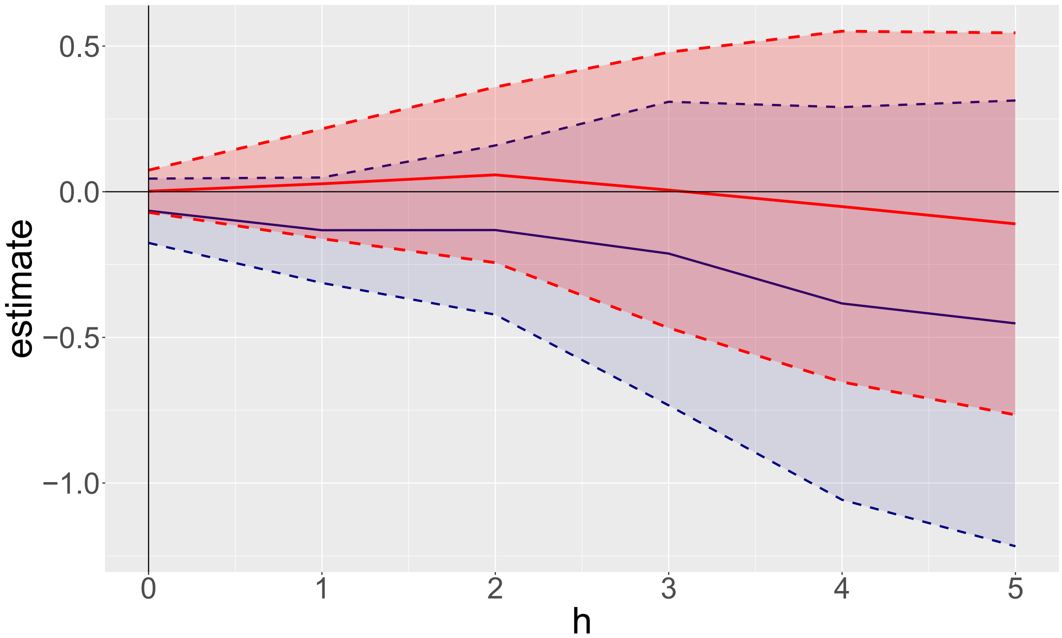
For product market reforms, the results are presented in table 5.9 and figure 5.13.

Table 5.9: Effects on employment: Product market reforms, narrative dataset

h	Estimate	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	0.002	0.037	0.045	0.964	747	0.406
1	0.027	0.096	0.284	0.776	747	0.352
2	0.058	0.154	0.377	0.706	747	0.241
3	0.006	0.241	0.024	0.981	747	0.157
4	-0.051	0.307	-0.166	0.868	747	0.117
5	-0.110	0.335	-0.330	0.742	747	0.092

The estimates show the effects of product market reforms on employment at h periods in the future, with the narrative dataset as the reform data.

Figure 5.13: Effects on employment: Product market reforms, narrative dataset



An impulse-response function comparison. The red curve represents the estimates with the narrative dataset, while the estimates with the data based on indicators are shown in blue. The hued regions represent the respective 95% confidence intervals.

For product market reforms, the effect is not statistically significant for any period. As compared to the previous results, the effect is much closer to zero. Surprisingly, slightly positive effects happen in the earlier periods. For years five and four, the effect is negative - although very small in absolute value.

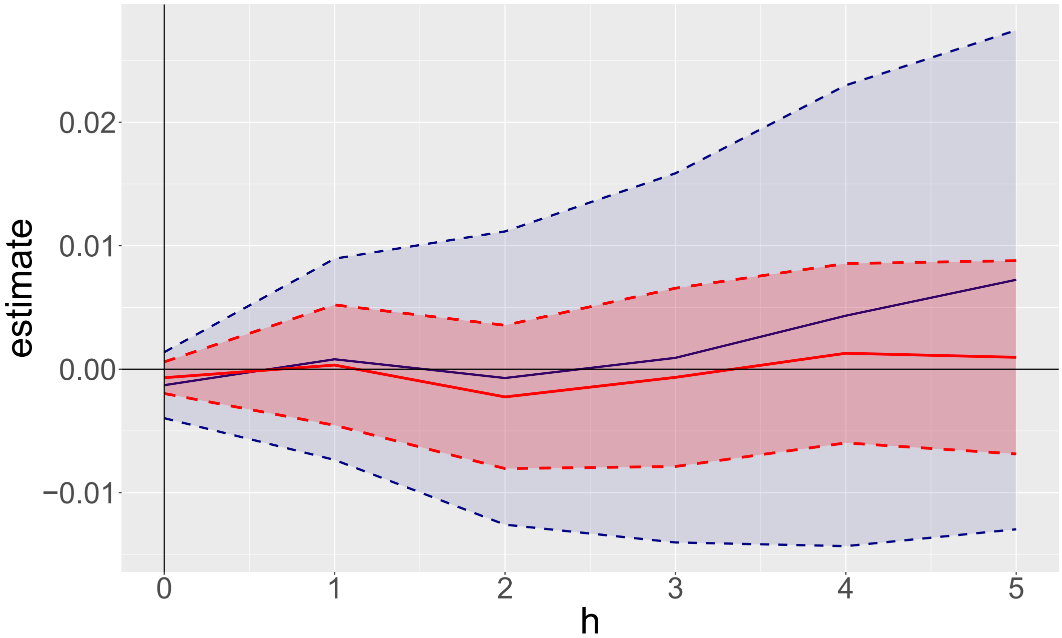
The effect of labour market reforms on GDP is then shown in table 5.10 and figure 5.14.

Table 5.10: Effects on GDP: Labour market reforms, narrative dataset

h	Estimate	$e^{Estimate}$	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	-0.001	0.999	0.001	-1.058	0.291	824	0.891
1	0.0003	1.0003	0.002	0.133	0.894	824	0.579
2	-0.002	0.998	0.003	-0.760	0.448	824	0.393
3	-0.001	0.999	0.004	-0.180	0.857	824	0.315
4	0.001	1.001	0.004	0.348	0.728	824	0.285
5	0.001	1.001	0.004	0.241	0.810	824	0.283

The estimates show the effects of labour market reforms on the logarithm of the GDP at h periods in the future, with the narrative dataset as the reform data.

Figure 5.14: Effects on GDP: Labour market reforms, narrative dataset



An impulse-response function comparison. The red curve represents the estimates with the narrative dataset, while the estimates with the data based on indicators are shown in blue. The hued regions represent the respective 95% confidence intervals.

The estimated effects of labour market reforms are not statistically significant. The effect sizes are very small. In fact, the more precise standard errors suggest a greater certainty of the effect being very close to zero, since for all periods, the hypothesis that the effect is $|\beta_{1,h}| \geq 0.01$ can be rejected for $h = 0, ..., 5$ (this is indicated by the confidence intervals).

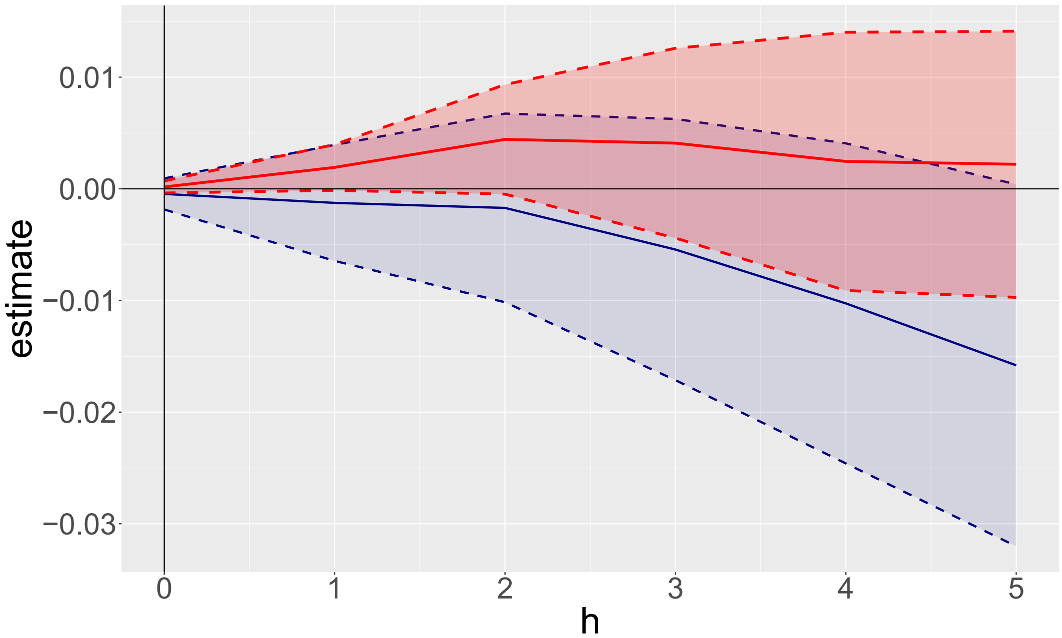
For the effects of product market reforms on the logarithm of the GDP, let us have a look at table 5.11 and figure 5.15.

Table 5.11: Effects on GDP: Product market reforms, narrative dataset

h	Estimate	$e^{Estimate}$	Standard error	t-stat.	p-value	N. obs.	Adjusted R^2
0	0.0002	1.0002	0.0003	0.609	0.543	824	0.891
1	0.002	1.002	0.001	1.835	0.067	824	0.580
2	0.004	1.004	0.003	1.769	0.077	824	0.396
3	0.004	1.004	0.004	0.944	0.346	824	0.316
4	0.002	1.002	0.006	0.416	0.677	824	0.285
5	0.002	1.002	0.006	0.362	0.718	824	0.282

The estimates show the effects of product market reforms on the logarithm of the GDP at h periods in the future, with the narrative dataset as the reform data.

Figure 5.15: Effects on GDP: Product market reforms, narrative dataset

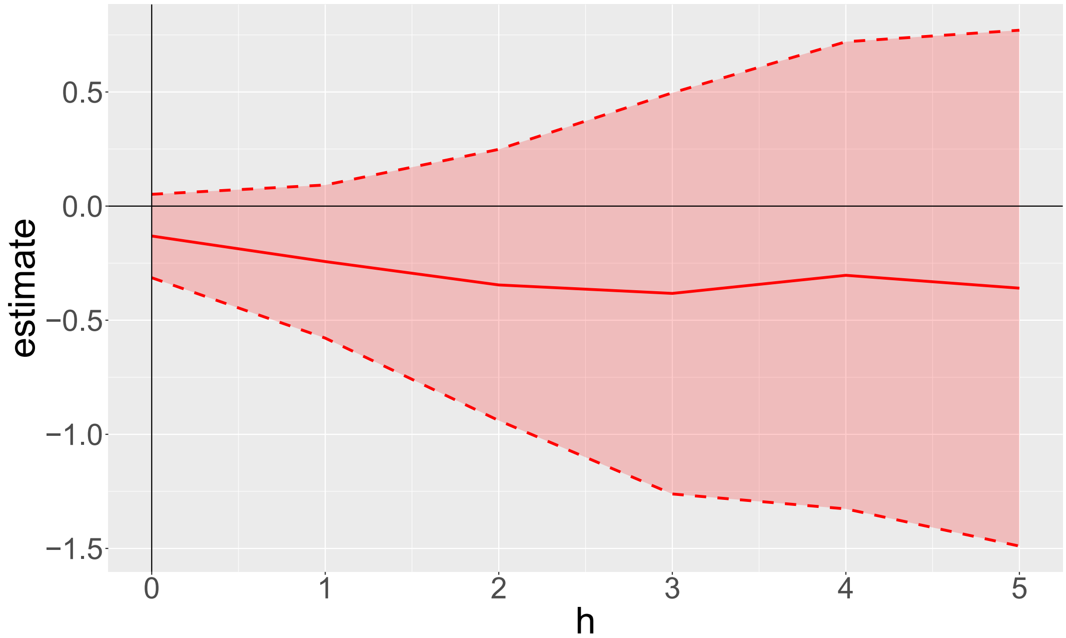


An impulse-response function comparison. The red curve represents the estimates with the narrative dataset, while the estimates with the data based on indicators are shown in blue. The hued regions represent the respective 95% confidence intervals.

The results are not statistically significant, except for periods one and two, where the results are significantly positive at the 90% level, with the multiplicative effect being a 0.2% to 0.4% increase in the GDP. Following these two periods however, the results are again far from being statistically significant and actually start to decline. Therefore, as opposed to the results in chapter 4, there is evidence for a slight positive effect on the economy in the short run, while the previous results showed more negative effects. However, the evidence is very limited.

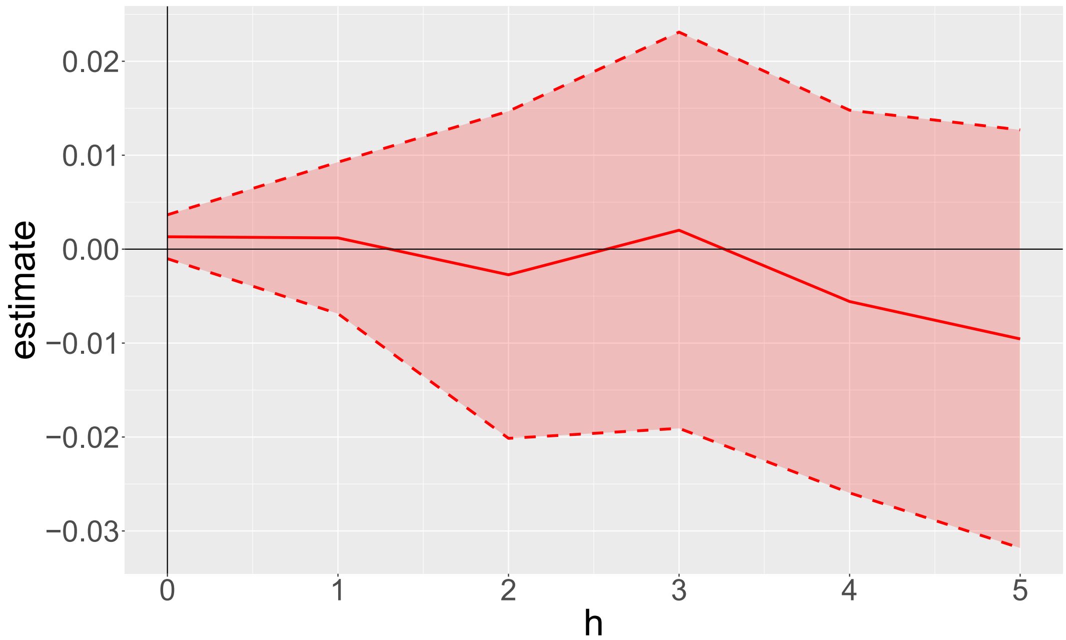
Lastly, the impulse response functions for negative labour market reforms (or reform reversals) are shown in figures 5.16 and 5.17. These were included in the labour market reform regressions as a separate dummy variable (see above). For the product market reforms, only two negative reforms have been detected in the whole dataset - therefore the results are not reported, since they would be practically irrelevant. The results for employment show a negative effect of the reversals, although the magnitude is quite small and not statistically significant. For the effects on the GDP, there is some fluctuation around zero followed by a decline. None of the effects are statistically significant either.

Figure 5.16: Negative labour market reforms: Effects on employment



An impulse-response function representing the effects of negative labour market reforms (regulation tightening) on employment.

Figure 5.17: Negative labour market reforms: Effects on GDP



An impulse-response function representing the effects of negative labour market reforms (regulation tightening) on the logarithm of the GDP.

In conclusion, the results show an increase in the statistical significance of the estimated effects as opposed to the previous results for certain estimations (labour - employment and product - GDP), while showing results fluctuating around zero more heavily for other estimations (product - employment and labour - GDP). All estimated effects are dramatically smaller than those of Bordon *et al.* (2016) and the hypothesis that structural reforms have important negative effects in the short run and important positive effects in the long run cannot be confirmed by the results in this section either.

Chapter 6

Conclusion

In this thesis, we have measured the dynamic effects of reforms on the economy. Specifically, we have concentrated on testing the robustness of a well-established model, authored by Bordon *et al.* (2016), measuring those effects. In the literature review, we have outlined some of the main issues and results found in the current literature on the topic. In the empirical estimation part of this thesis, we have tested the influence of an extended dataset on the effects of reforms, as well as that of changing the measurement of several variables. Then, to confirm whether we could obtain similar results to the Bordon *et al.* (2016) paper, we have restricted the dataset. Lastly, robustness check have been performed. These included a change in the estimation of the output gap and the replacement of the reform data with a narrative dataset.

In general, we have found quite different results from the established literature. In our estimations, reforms seem to have smaller effects, which are not statistically significant most of the time. With respect to Bordon *et al.* (2016), the dataset extension has shown that withdrawing countries from the sample without proper justification can have important effects on the results. Furthermore, the dataset extension in the time dimension might have had a noticeable effect on the results as well - namely in reducing the dynamic bias of the estimation. However, it is also possible that the difference in the results might not be just arising from the methodological differences. There is a possibility that there have been, in recent years, important changes in the structure of the OECD economies that have also had an influence on the effects of reforms, such as the financial crisis of 2008. The results therefore raise doubt on the seemingly answered and clear question of what the effects of reform are. We have not been able to confirm the usual hypothesis that structural reforms have neutral or negative effects in the short run and positive effects in the long run.

Further, our robustness checks results indicate that more attention should be paid to the way the output gap is measured. Since the results have been noticeably affected by changing the output gap measurement method, we suggest that the incorrect choice of measurement of this variable can bring substantial bias in the results. Yet this seems to be scarcely addressed in the literature. Moreover, changing the way reforms are measured has also had drastic effects on the results. This issue is treated in the literature often. However, a satisfactory and widely accepted way of measuring reforms is yet to appear. Since the change in the results is so dramatic, research should perhaps dig even deeper into this issue.

Therefore, while it might seem that our research question has been answered by previous research, the instability and the non-robust nature of the results suggests otherwise. A question might even be whether panel data macroeconometric estimations are suited to answer the question at hand at all. Other methods might capture the heterogeneity of the effects of reforms across countries better. Indeed, there have been hints of other methods in the research that might complement the view based on the local projections, specifically the synthetic control method and case studies (Adhikari *et al.* (2016), Marrazzo & Terzi (2017)).

In conclusion, it needs to be said that it is important not to settle with the established results. The answer to the question of the effects of structural reforms is important in practice in economic policy making. If their positive effects are taken for granted, structural reforms might be seen as a universal cure. Indeed, structural reforms are often undertaken in times of crisis rather than expansion (Dias da Silva *et al.* 2017), possibly being seen as a way out of the crisis. However, this approach might be unfounded. Lastly, the well-being of a society cannot be measured purely by macroeconomic indicators such as the GDP or employment. More attention should be also paid to the social effects of reforms. While there might be positive effects on the whole economy, the reforms (such as the weakening of job protection) might hit hard on some members of society. Therefore, more research should also focus on these issues.

References

ABIAD, A., E. DETRAGIACHE, & T. TRESSEL (2008): "A New Database of Financial Reforms." *IMF Working Paper WP/08/266*, International Monetary Fund.

ADHIKARI, B., R. DUVAL, B. HU, & P. LOUNGANI (2016): "Can Reform Waves Turn the Tide? Some Case Studies Using the Synthetic Control Method." *IMF Working Paper WP/16/171*, International Monetary Fund.

ALESINA, A., D. FURCERI, J. D. OSTRY, C. PAPAGEORGIOU, & D. QUINN (2020): "Structural Reforms and Election: Evidence from a World-Wide New Dataset." *Journal of Economic Literature* .

DE ALMEIDA, L. A. & V. BALASUNDHARAM (2018): "On the Impact of Structural Reforms on Output and Employment: Evidence from a Cross-Country Firm-Level Analysis." *IMF Working Paper WP/18/73*, International Monetary Fund.

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 (2013): "Structural Reforms and Economic Growth: A Meta-Analysis." *Working Paper Series 8*, Czech National Bank.

DE BANDT, O., O. VIGNA *et al.* (2008): "The macroeconomic impact of structural reforms." *Quarterly Selection of Articles 11*, Banque de France.

BASSANINI, A. (2015): "A Bitter Medicine? Short-term Employment Impact of Deregulation in Network Industries." *IZA Discussion Papers 9187*, Institute of Labor Economics.

BERG, A., J. D. OSTRY, & J. ZETTELMEYERB (2012): "What makes growth sustained?" *Journal of Development Economics* **98**: pp. 149 – 166.

Bordon, A. R., C. Ebeke, & K. Shirono (2016): "When Do Structural Reforms Work? On the Role of the Business Cycle and Macroeconomic Policies." *IMF Working Paper WP/16/62*, International Monetary Fund.

Bouis, R., O. Causa, L. Demmou, & R. Duval (2012a): "How quickly does structural reform pay off? An empirical analysis of the short-term effects of unemployment benefit reform." *IZA Journal of Labor Policy* .

Bouis, R., O. Causa, L. Demmou, R. Duval, & A. Zdzienicka (2012b): "The Short-Term Effects of Structural Reforms: An Empirical Analysis." *OECD Economics Department Working Papers 949*, Organisation for Economic Co-operation and Development.

Christiansen, L., M. Schindler, & T. Tressel (2009): "Growth and Structural Reforms: A New Assessment." *IMF Working Paper WP/09/284*, International Monetary Fund.

Croissant, Y., G. Millo, K. Tappe, O. Toomet, C. Kleiber, A. Zeileis, A. Henningsen, L. Andronic, & N. Schoenfelder (2020): *plm: Linear Models for Panel Data.* R package version 2.2-3.

Dabla-Norris, E., G. Ho, & A. Kyobe (2016): "Structural Reforms and Productivity Growth in Emerging Market and Developing Economies." *IMF Working Paper WP/16/15*, International Monetary Fund.

De Grauwe, P., Y. Ji *et al.* (2016): "Flexibility versus Stability: a difficult trade-off in the eurozone (Appendix II)." *CEPS Working Document 422*, CEPS.

De Masi, P. *et al.* (1997): "IMF Estimates of Potential Output; Theory and Practice." *Technical report*, International Monetary Fund.

Driscoll, J. C. & A. C. Kraay (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *The Review of Economics and Statistics* **80(4)**: pp. 549–560.

Duval, R. & D. Furceri (2018): "The effects of labor and product market reforms: the role of macroeconomic conditions and policies." *IMF Economic Review* **66(1)**: pp. 31–69.

Duval, R., D. Furceri, B. Hu, J. Jalles, & H. Nguyen (2018): "A Narrative Database of Major Labor and Product Market Reforms in Advanced Economies." *IMF Working Paper WP/18/19*, International Monetary Fund.

Duval, R., D. Furceri, & J. Jalles1 (2017): "Job Protection Deregulation in Good and Bad Times." *IMF Working Paper WP/17/277*, International Monetary Fund.

Egert, B. (2017): "Regulation, institutions and productivity: New macroeconomic evidence from OECD countries." *OECD Economics Department Working Papers 1393*, Organisation for Economic Co-operation and Development.

Gal, P. N. & A. Hijzen (2016): "The short-term impact of product market reforms: A cross-country firm-level analysis." *IMF Working Paper WP/16/116*, International Monetary Fund.

Hamermesh, D. S. (2007): "Replication in economics." *Canadian Journal of Economics/Revue canadienne d'économique* **40(3)**: pp. 715–733.

Hamilton, J. D. (2018): "Why you should never use the Hodrick-Prescott filter." *Review of Economics and Statistics* **100(5)**: pp. 831–843.

Hartung, J. (1998): "A note on combining dependent tests of significance." *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **41(7)**: pp. 849–855.

Hastie, T. & R. Tibshirani (1986): "Generalized Additive Models." *Statistical Science* **1(3)**: pp. 297–318.

Herbst, E. P. & B. K. Johannsen (2020): "Bias in Local Projections." *Finance and economics discussion series (feds)*, Board of Governors of the Federal Reserve System.

Hodrick, R. J. & E. C. Prescott (1997): "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking, Vol. 29, No. 1* **29(1)**: pp. 1–16.

IMF (2016): *Chapter 3. Time for a Supply-Side Boost? Macroeconomic Effects of Labor and Product Market Reforms in Advanced Economies*, pp. 110–111. International Monetary Fund.

IMF (2019): "World Economic Outlook Database." https://www.imf.org/external/pubs/ft/weo/2019/01/weodata/index.aspx. Accessed: 22/11/2019.

Jorda, O. (2005): "Estimation and inference of impulse responses by local projections." *American economic review* **95(1)**: pp. 161–182.

Kleiber, C. & C. Lupi (2011): "Panel unit root testing with R." *R-Forge, R-Project.* .

Koske, I., I. Wanner, R. Bitetti, & O. Barbiero (2015): "The 2013 update of the OECD's database on product market regulation: Policy insights for OECD and non-OECD countries." *OECD Economics Department Working Papers 1200*, Organisation for Economic Co-operation and Development.

Laeven, L. & F. Valencia (2013): "Systemic banking crises database." *IMF Economic Review* **61(2)**: pp. 225–270.

Laeven, L. & F. Valencia (2018): "Systemic banking crises revisited." *Working paper WP/18/206*, IMF.

Marrazzo, P. M. & A. Terzi (2017): "Wide-reaching structural reforms and growth: A cross-country synthetic control approach." *CID Research Fellow and Graduate Student Working Paper Series* .

OECD (2013): "Protecting jobs, enhancing flexibility: A new look at employment protection legislation." *OECD Employment Outlook 2013* pp. 65 – 126.

OECD (2014): "Quarterly National Accounts." https://www.oecd-ilibrary.org/content/data/data-00017-en. Accessed: 12/01/2020.

OECD (2018): "Indicators of Product Market Regulation." https://www.oecd.org/economy/reform/indicators-of-product-market-regulation/. Accessed: 11/02/2020.

Okun, A. (1963): "Potential GNP: its measurement and significance." *Cowles Foundation Paper* **190**.

Piton, C. & F. Rycx (2018): "The Unemployment Impact of Product and Labour Market Regulation: Evidence from European Countries." *IZA Discussion Papers 11582*, Institute of Labor Economics.

Plagborg-Møller, M. & C. K. Wolf (2019): "Local projections and VARs estimate the same impulse responses." *Unpublished paper: Department of Economics, Princeton University* p. 1.

Romer, C. D. & D. H. Romer (1989): "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual* **4**: pp. 121–170.

Dias da Silva, A., A. Givone, & D. Sondermann (2017): "When Do Countries Implement Structural Reforms?" *ECB Working Paper 2078*, European Central Bank.

Teulings, C. N. & N. Zubanov (2010): "Is Economic Recovery a Myth? Robust Estimation of Impulse Responses." *CESIFO WORKING PAPER* **(3027)**.

Tola, A. & S. Waelti (2018): "Financial crises, output losses, and the role of structural reforms." *Economic Inquiry* **56(2)**: pp. 761–798.

Wooldridge, J. M. (2002): *Econometric analysis of cross section and panel data.* MIT press.

Appendix A

A.1 Lagged dependent variable specifications

In terms of the model specification, equation (3.1) differs from that of Bordon *et al.* (2016) by not including the following term: $\psi_h(L)dep\_var_{i,t+h-1}$, where $dep\_var_{i,t+h} = y_{i,t+h} - y_{i,t-1}$. It is not quite clear what the authors meant by this. They do not offer any clarification in the text besides this expression in the equation. The standard definition of $a(L)x_t$ is a lag polynomial: $a_0x_t + a_1x_{t-1} + a_1x_{t-2} + ... + a_px_{t-p}$ for some p. However, how exactly should this be applied to this case? What should be p equal to? Three interpretations of this term have been tried:

1. $\psi_h(L)dep\_var_{i,t+h-1} = a_1(e_{i,t+h-1} - e_{i,t-1}) + \ldots + a_{h-1}(e_{i,t} - e_{i,t-1})$
   The (L) would here indeed function as a sort-of a lag polynomial operator.

2. $\psi_h(L)dep\_var_{i,t+h-1} = a_1(e_{i,t+h-1} - e_{i,t-1})$
   Here, the (L) is taken to mean a simple lag in the h period.

3. $\psi_h(L)dep\_var_{i,t+h-1} = a_1(e_{i,t+h-2} - e_{i,t-2})$
   Lastly, (L) is taken to mean as an additional lag to both periods.

However, none of them have given quite meaningful results (see the results below with 90% and 95% confidence intervals - the underlying dataset is restricted as in the first part of the robustness checks section), being quite far from those of Bordon *et al.* (2016). The authors probably used a different specification which we were not able to decode. However, no similar specification seems to exist in the literature on structural reforms (at least for the papers included in the literature review). There is also no apparent reason that such a specification would reduce the omitted variable bias. Therefore, this term has been omitted from the equations estimated in the thesis.

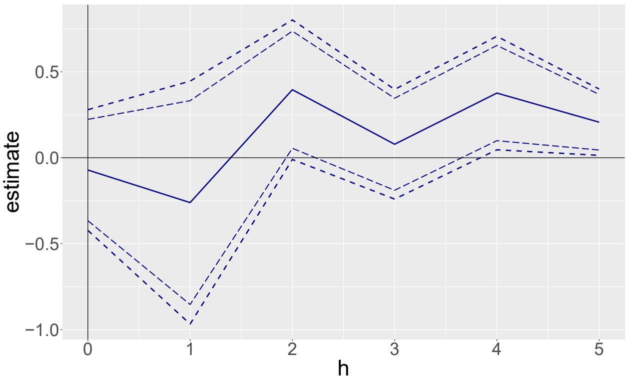


Figure A.1: Labour reforms, specification 1

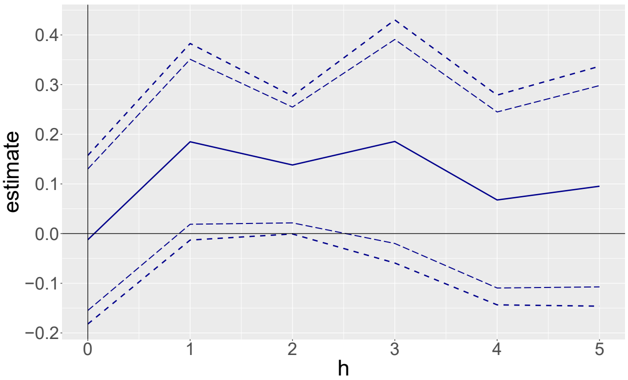


Figure A.2: Product reforms, specification 1

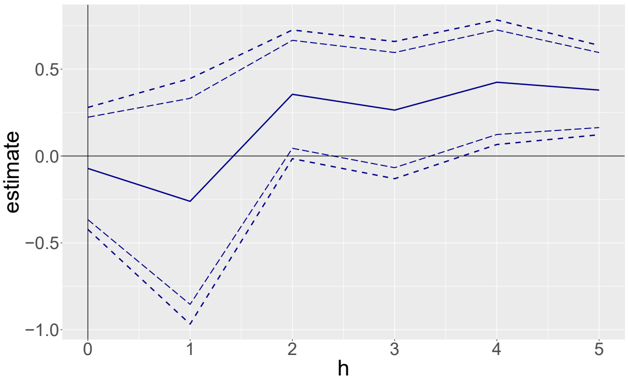


Figure A.3: Labour reforms, specification 2

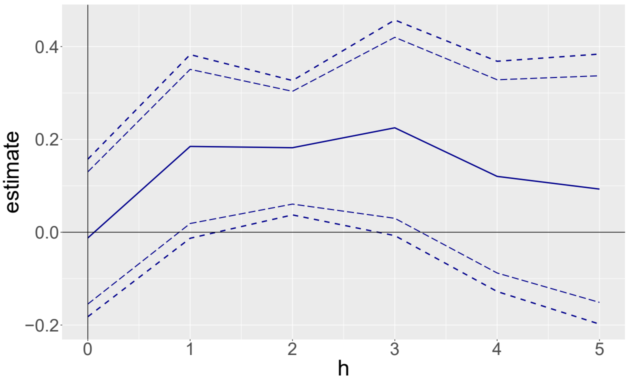


Figure A.4: Product reforms, specification 2

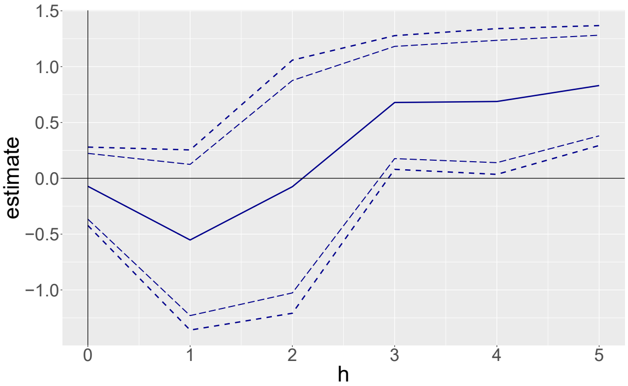


Figure A.5: Labour reforms, specification 3

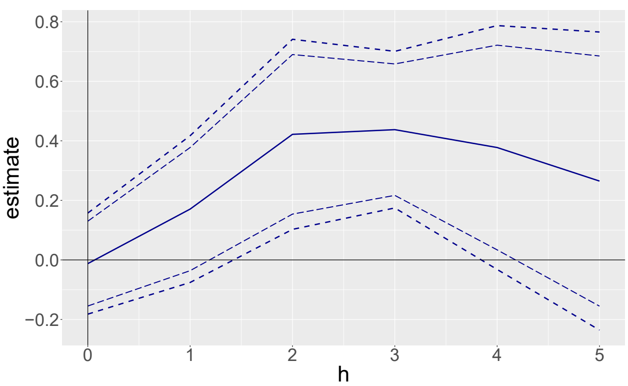


Figure A.6: Product reforms, specification 3