Essays on Fiscal Policy

Marcell Göttert

Dissertation

Essays on Fiscal Policy

Inauguraldissertation
zur Erlangung des akademischen Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Volkswirtschaftlichen Fakultät
an der Ludwig-Maximilians-Universität München

2022

vorgelegt von
Marcell Göttert

Referent: Prof. Dr. Andreas Peichl

Korreferent: Prof. Dr. Jesús Crespo Cuaresma

Tag der mündlichen Prüfung: 04. Juli 2022

Promotionsabschlussberatung: 20. Juli 2022

To my grandparents, Elvi and Hubert Göttert

Acknowledgements

First, I thank my supervisor Andreas Peichl for his support, advice and encouragement regarding this thesis and my steps as an economist in my time at the ifo as well as the Agenda Austria. Moreover, I would like to thank Jesús Crespo Cuaresma for agreeing to be my co-supervisor and many helpful comments regarding the last chapter of this thesis. I also thank Andreas Haufler for completing the thesis committee.

Furthermore, I thank my co-author Timo Wollmershäuser for helpful comments and support on the joint research project and for enhancing my knowledge on business cycle analysis and forecasts. Moreover, I am indebted to Robert Lehmann for insightful comments and support on the joint research project and teaching me many things about research and policy projects. I also sincerely thank Christian Grimme for his helpful comments especially on the last chapter of this thesis and the advices on conducting research or policy consulting projects.

I am also indebted to many colleagues at the ifo Institute as well as the Agenda Austria for providing me with an inspiring and pleasant work environment. I have to thank many advisers and colleagues, in particular Maximilian Blömer, Carla Krolage, Denés Kucsera, Stefan Lautenbacher, Heike Lehner, Sebastian Link, Hanno Lorenz, Sascha Möhrle, Magnus Reif, Gerhard Reitschuler, Franz Schellhorn, Felix Schröter and many others.

I have to thank Steffen Therre, one of my oldest friends, for proofreading this thesis and providing me with a detailed feedback. Thank you also for sharing many years of friendship and the same kind of humor.

Moreover, I thank my family and friends for their support in my entire life. I am especially grateful to my parents Christine and Horst as well as my sister Saskia.

Last but not least, I am indebted to my late grandparents Elvi and Hubert for the many talks, for supporting me, for teaching me so much and never having any doubt that I would successfully complete this thesis.

Contents

List of Figures

List of Tables

Introduction

Pecunia non olet – Money does not stink. Who does not recognize the famous saying, which is ascribed to the Roman emperor Vespasian? He is said to have spoken those words in an argument regarding the taxation of public toilets in ancient Rome almost 2000 years ago. The quote is a relict of one of the oldest arguments about fiscal policy. The importance and the understanding of fiscal policy increased substantially ever since. Almost 15 years ago, the Great Recession hit the world and most of the governments reacted with large spending programs. The recent Corona crisis led to even larger spending programs. The United States of America responded to COVID-19 with a stimulus that amounts to more than four trillion Dollars. Similarly, the European Union reacted with the NextGenerationEU fund that has an amount of more than two trillion Euro. It is the largest stimulus package in the history of the European Union. Fiscal policy plays an important role in our society and we should try to enhance our knowledge about its tools, measures and effects as much as possible.

This thesis tries to shed some light on three important aspects regarding fiscal policy. It consists of three essays on measuring tools for the current fiscal rules in the European Union, on tax forecasts and on the effects of social security changes.

The first chapter *Survey-Based Structural Budget Balances* analyzes the main properties of the current methods on calculating structural budget balances. It is joint work with Timo Wollmershäuser. We find large revisions for the production function based approaches used by the European Commission the International Monetary Fund and the Organisation for Economic Co-operation and Development. Moreover, we can reveal that their method results in a systematic overestimation of the ex-post structural budget balance via the output gap in the investigated period. Applying a Hodrick-Prescott Filter was the main method for constructing structural budget balances before the most important institutions switched to a production function-approach. Our analysis shows that moving back to a Hodrick-Prescott filter would not solve the problem. Instead, we construct the capacity utilization gap as a firm survey-based replacement for the production function-based output gap. The newly constructed structural budget balances are highly correlated with the existing ones. However, the revision size is substantially smaller and the bias is nearly nonexistent. The survey-based method can be easily implemented into the existing framework of EU fiscal rules and many existing national fiscal rules.

The second chapter *Tax Revenue Forecast Errors: Wrong Predictions of the Tax Base or the Elasticity?*, which is joint work with Robert Lehmann, tries to determine the source of tax forecast errors in Germany. We investigate the tax forecasts for the six largest tax types and the overall tax sum for Germany by the Working Party on Tax Revenue Forecasting (*Arbeitskreis Steuerschätzungen*). Our results reveal that the estimates for some tax types are biased. The Working Party uses the macroeconomic projection by the Federal Ministry for Economic Affairs and Energy. We found, that e.g. their forecast for the nominal gross domestic is too optimistic in the investigated period. When looking for the source of the tax forecast errors, we find mixed results. The forecast errors of the profit-related taxes and the wage tax are mainly driven by the errors of the macroeconomic tax base. For the energy tax and the sales taxes, the wrong predictions of the elasticity matter more. More than two-thirds of the forecast error of the overall tax sum can be attributed to false macroeconomic predictions and roughly one-third to wrong assumptions on the elasticity.

The third chapter *The State-Dependent Effects of Social Security Contributions on the Macroeconomy* focuses on the effect of changes in social security contributions on various macroeconomic variables and if the effects differ depending on the position in the business cycle. I follow a narrative approach and exploit OECD forecast errors for Canada, Japan, the United Kingdom and the United States of America between 2004 and 2018. The macroeconomic effects are estimated via Local Projections in a panel setting. I find differences in the response to a shock in social security contributions when comparing recessions with non-recessionary periods. The effects on the real gross domestic product follow a pro-cyclical pattern. They are mainly driven by the stimulation of real private consumption and real capital formation. The pro-cyclical effects found in this chapter fits to the response in the existing literature regarding tax multipliers, which are also found to follow such a pattern. My results suggest a cut in social security contributions might not be the best choice when considering components for a stimulus package to counter an economic downturn.

The appendix and the consolidated bibliography can be found at the end of this thesis.

Chapter 1

Survey-Based Structural Budget Balances [1]

1.1 Introduction

One of the issues of the dispute between Italy and the European Commission (EC) in 2018 about the Italian budget and its compliance with the fiscal rules of the European Union (EU) is the calculation of the structural budget balance. At the core of this dispute is the estimation of potential output, which currently follows a production function-based approach prescribed by the EC (Roeger *et al.*, 2019).[2] This approach attracted widespread criticism, both for the large ex post revisions of the resulting output gap estimates (e.g. Marcellino and Musso, 2011; Kempkes, 2014) and the underlying estimation of the structural unemployment rate (e.g. Fioramanti and Waldmann, 2016). Given this evidence Tooze (2019) launched a campaign against “nonsense output gaps” by arguing that, “when combined with stringent fiscal rules, backward-looking estimates of potential output can have truly perverse effects.”

In this paper we investigate the structural budget balances published by the EC, the Organisation for Economic Co-operation and Development (OECD) and the International Monetary Fund (IMF). The three institutions give economic policy recommendations to their member countries and analyze their economic situation. These recommendations and analyses also refer to fiscal policy and include fiscal surveillance. We show that the structural budget balances of these institutions, which are all estimated on the basis of a production function, are

[1] This chapter is joint work with Timo Wollmershäuser. An older version circulates as Göttert and Wollmershäuser (2021).

[2] Ademmer and Dreher (2016) discuss the political budget cycles in the EU and the influence of fiscal rules and the media strength on them.

indeed systematically revised downward after their initial release. The source of this bias is the underlying output gap estimation which suffers from decomposing the level of GDP into a stable trend and cycle. Thus, critics are right when they argue that the production function-based approach of estimating structural budget balances is an inappropriate tool for fiscal surveillance.

Therefore, we propose a survey-based approach that avoids the shortcomings of the production function-based approach. Our measure is based on the degree of capacity utilization of firms, which is calculated from a representative business survey. It is available without publication lag in the middle of the current quarter and is highly correlated with the existing estimates for the output gaps. We show that in comparison with the production function-based approach structural budget balances calculated using survey-based capacity utilization are significantly less biased and hardly revised. This high reliability, precision and early availability make our new estimation method a perfect tool for fiscal surveillance and the EU fiscal framework. It could be easily implemented into the existing EU fiscal rules without any major changes and be interpreted as an improvement of the design of the fiscal rules in the EU. In the end, the effectiveness of the disciplinary tools of the EU fiscal framework could be improved, which according to Caselli and Reynaud (2020), DeJong and Gilbert (2020) and Bergmann *et al.* (2016) contributes to the reduction of fiscal deficits. A detailed discussion of the existing EU fiscal rules as in Buti and Carnot (2012) is not the aim of this paper. However, we contribute to the discussion of the surveillance of the EU fiscal rules.

There are a number of papers on the evaluation and the construction of output gaps. Some of them focus on the use of capacity utilization in the manufacturing sector for improving output gap estimates, but most of them are silent regarding the consequences of output gap revisions on structural budget balances. Kempkes (2014) evaluates the output gaps published by the EC, the OECD and the IMF. His analysis concentrates on the revisions of output gap forecasts, but ignores revisions of ex post output gaps. Moreover, he argues that the biased output gap forecasts can lead to significant additional public debt. Marcellino and Musso (2011) show that the results of most of the output gap estimation methods for the euro area are very uncertain and unstable over time. They recommend the use of capacity utilization to determine the output gap and attest it both, high stability and low uncertainty. Orphanides *et al.* (2000) show the high correlation between output gap estimates and capacity utilization. Nyman (2010), Graff and Sturm (2012), Hulej and Grabek (2015), Szörfi (2015) and Silva *et al.* (2016) include capacity utilization or other survey data as cyclical information into their empirical models for decomposing the level of GDP into trend and cycle. Their survey-based measures have smaller revisions of the cyclical component than the traditional methods. In this paper we follow the recommendation of Marcellino and Musso (2011) and directly use capacity utilization as a measure for the output gap. We construct an aggregate capacity utilization by combining survey data of the manufacturing and the service sector and apply it to the estimation of ex post structural budget balances.

The remainder of this paper is organized as follows. Section 1.2 gives a short introduction to the concept of structural budget balances. Sections 1.3 and 1.4 evaluate the mainstream methods of estimating structural budget balances. Section 1.5 introduces a survey-based method and compares it to the mainstream approach. Section 1.6 concludes.

1.2 The Structural Budget Balance

Structural budget balances are used to evaluate the budget balance without the distorting influences of the business cycle and temporary one-offs. They are mostly used for fiscal surveillance and of utmost importance for the EU fiscal rules. One of the main parts of the fiscal framework in the EU is laid out by the Stability and Growth Pact (SGP), which consists of two branches: a preventive and a corrective arm. While the preventive arm ensures fiscal stability by preventing an excessive deficit and excessive debt, the corrective arm helps a member state out of an excessive deficit or excessive debt. In the preventive arm, some of the spending goals are formulated in terms of the structural budget balance and evaluated ex post. In the corrective arm, the path out of an excessive deficit or debt is a combination of an expenditure path and a structural budget balance path, which is also evaluated ex post. Moreover, the requirements for the expenditure path and consequently the reduction of the debt level depend on the output gap (ECOFIN, 2017; EC, 2018). However, the fiscal policy of the EU member states is not only regulated by the SGP but also by the fiscal compact. There, it is set out the signatory states may not have a structural deficit larger than 0.5% of gross domestic product (EU, 2012).[3]

Technically, the structural budget balance (X) is computed by subtracting one-offs (T) and the influence of the business cycle from the budget balance (B). The cyclical component is calculated as the product of the measure of the business cycle (Y) and the semi-elasticity of the budget balance with respect to the business cycle ($\epsilon_{B,Y}$):

$$X = B - \epsilon_{B,Y} Y - T. \tag{1.1}$$

In the approach used by the EC, the OECD and the IMF the business cycle Y is measured by the output gap (GAP), which is defined as the percentage deviation

[3]Under the most favorable circumstances signatory states may have a deficit up to 1% of gross domestic product.

of real gross domestic product (GDP) from potential output (POT):

$$GAP = \frac{GDP - POT}{POT} \cdot 100\%. \tag{1.2}$$

The major shortcoming of this approach is the estimation of the output gap. Unlike GDP, "potential output is, and always will be, an unobservable variable and consequently has to be estimated" (Buti *et al.*, 2019). So, there is nothing like the "true" output gap, but various methods of how to estimate it. These methods can be evaluated depending on their use. For calculating structural budget balances, which are important for fiscal surveillance and policy advice, it is crucial that the estimation method produces unbiased results with little revisions. Furthermore, it would be desirable to have an approach that can be easily explained to political institutions and the public, that is free of political influence and that produces timely results.

The methods for calculating the output gap can be divided into a time series approach and a production function approach. In the time series approach the series is split with a filtering method into trend and cycle. The most commonly used method is the Hodrick-Prescott filter (Hodrick and Prescott, 1997). The major shortcoming of this filter is the so-called endpoint problem. Changing values at the end of the series, due to e.g. data revisions or new data coming in, often turns out to have a large impact on the estimation of the trend component (Mise *et al.*, 2005). In practice, attempts are being made to mitigate the endpoint problem by extending the series at the end with forecast values (Kaiser and Maravall, 2001). However, as these forecasts are subject to even larger changes than the ex post data, the estimates of the trend component in real-time are still frequently revised. Other filter techniques, such as a band-pass filter are suffering from the same problem (Orphanides and van Norden, 2002; Cayen and van Norden, 2005; Marcellino and Musso, 2011).

The production function approach of calculating an output gap defines a production function that determines POT. Again, the output gap is calculated as a residual. Most of the trend components of the arguments of the production function are derived via a filtering method, which is quite often a Hodrick-Prescott filter. Thus, the production function approach is subject to the same problem as the time series approach and only shifts it from the level of GDP to the level of the arguments of the production function. Similar to the time series approach, the production functions are usually calculated not only for ex post, but also for future values and hence, include forecasts. Again, changes in the forecasts lead to changes in the estimated trend and cycle component. Thus, like the time series approach the production function approach is also subject to large revisions.

1.3 Revisions of the Structural Budget Balance: Production Function Approach

1.3.1 Data and Revision Size

For the evaluation of the production function approach we use data releases of the EC, the OECD and the IMF for Germany (DE), France (FR), United Kingdom (UK), Italy (IT), Spain (ES) and Austria (AT). Each institution has two data releases per year, one in spring and one in autumn. For each country and each institution we use the vintages ranging from spring 2003 until spring 2018. The real-time vintages of the EC are downloaded from the EC's Communication and Information Resource Centre for Administrations, Businesses and Citizens (CIRCABC). The real time data of the OECD and the IMF has been extracted from their biannual analyses, the OECD Economic Outlook and the IMF World Economic Outlook.[4] For every institution, we use their country-specific semi-elasticity to calculate the structural budget balance. The semi-elasticities are available in Mourre *et al.* (2019) for the EC, Girouard and André (2005) for the OECD and IMF (1993) for the IMF. [5]

We focus our analysis on ex post revisions of the structural budget balance. We denote by $X_{k,l,t+i}$ the ith ex post estimation of the structural budget balance of institution k for country l and year t. The first ex post estimation for year t is released in spring of the year following t (i.e. i=1), the second in autumn of the year following t (i.e. i=2), and so on. We assume that ex post revisions of the structural budget balance are only driven by revisions of the cyclical component and hence the estimation of the output gap. Thus, ex post data revisions of the budget balance B or the one-offs T, or revisions of the semi-elasticity $\epsilon_{B,GAP}$ are excluded from the analysis.[6] For $j > i$ the ex post revisions of the structural budget balance are then given by

$$X_{k,l,t+i} - X_{k,l,t+j} = \epsilon_{B,GAP,k,l}(GAP_{k,l,t+i} - GAP_{k,l,t+j}). \tag{1.3}$$

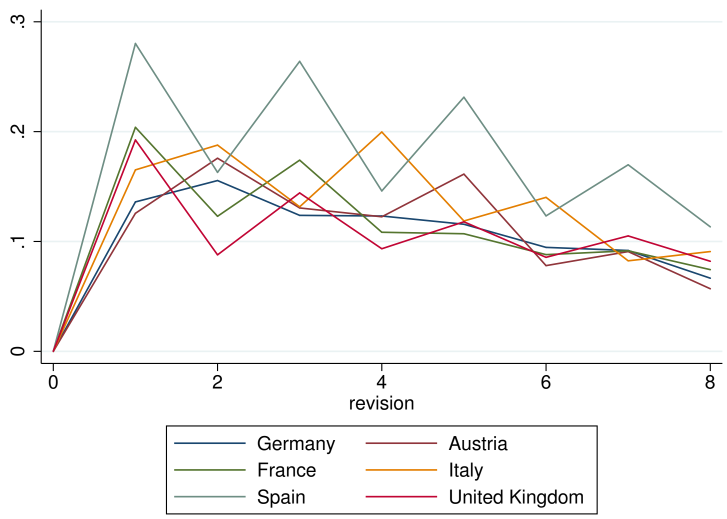
[4]The commonly agreed method in the EU, which is used by the EC, is described in Havik *et al.* (2014). The method used by the OECD is explained in OECD (2012). The IMF has no commonly agreed method, but uses country-specific approaches as argued in deResende (2014).

[5]The IMF did not publish a semi-elasticity for Spain and Austria in IMF (1993). We used the semi-elasticities of the EC as a replacement. Thus, a different semi-elasticity means a different level of the structural budget balance. However, if the semi-elasticity remains unchanged over time, it does not induce revisions. Consequently, it cannot be the source of a possible bias.

[6]The EC changed its semi-elasticities for the member countries only slightly in the last years (Mourre *et al.*, 2019). To our knowledge, the OECD and the IMF left their semi-elasticities unchanged over the years. In any case, new estimations of the underlying elasticities for the subcomponents lead to pretty similar results over the years (Price *et al.*, 2014).

Figure 1.1 shows the size of the revisions, which is calculated for each country as the mean over all years t ranging from 2002 to 2017 and all institutions k of the absolute ex post revision of the structural budget balances between two subsequent vintages $t+i$ and $t+i+1$. The size of the revisions is sizable and persistent. It is 0.28 percentage points of potential output for the first revision for Spain, 0.20 for France, 0.19 for the UK, 0.17 for Italy, 0.14 for Germany and 0.13 for Austria. The revision size declines only slowly for later revisions. For the eighth revision (autumn release four years after t compared to the spring release four years after t) it is still between 0.06 and 0.11 percentage points of potential output. The largest revisions are found for Spain. To some extent these revisions are caused by revisions of ex post GDP data, but the largest share is due to forecast errors and forecast revisions that are used to mitigate the endpoint problem.[7]

Figure 1.1: Size of the ex post revisions of the structural budget balance between two subsequent vintages using the production function approach (in percentage points of potential output)



1.3.2 Biasedness

Large revisions of ex post structural budget balances make the application and implementation of EU fiscal rules difficult and the recommendations based on them uncertain and contestable. However, the problem becomes even bigger, if the

[7]In Appendix A.1 we show that the relative size of the revisions of the cyclical component is multiple of the relative revision of GDP, nominal GDP or the budget balance (relative to nominal GDP). However, we also show that the influence of the revisions of the budget balance on the structural budget balance is also non-negligible.

revisions not only turn out to be large, but systematic and biased. To investigate this we follow Kempkes (2014) and apply an empirical model that was developed by Holden and Peel (1990a):

$$X_{l,t+1} - X_{l,t+j} = \alpha_{l,j} + u_{l,j}, \tag{1.4}$$

where for each country l the revision of ex post structural budget balances between the vintage j and their initial estimate $i = 1$ is regressed on a constant α. For the estimation we pooled the data of the three institutions and applied Newey-West standard errors.

In an ideal world, the estimation method for the structural budget balance should on average correctly assesses the cyclical component of the budget balance already with the initial estimation. In such a case α in equation (1.4) would be equal zero. If, however, α turns out to be significantly different from zero, the estimation method for the structural budget balance is biased. If $\alpha < 0$, the structural budget balance X initially turns out to be overestimated and is revised downward in later estimations. These revisions can be traced back to a systematic underestimation of the output gap and consequently the cyclical component of the budget balance. If $\alpha > 0$, the opposite is the case. The structural budget balance X would be systematically revised upward. The output gap would initially be overestimated and the structural budget balance underestimated. So, the structural budget balance in the initial estimation would be systematically lower than in the later ones.

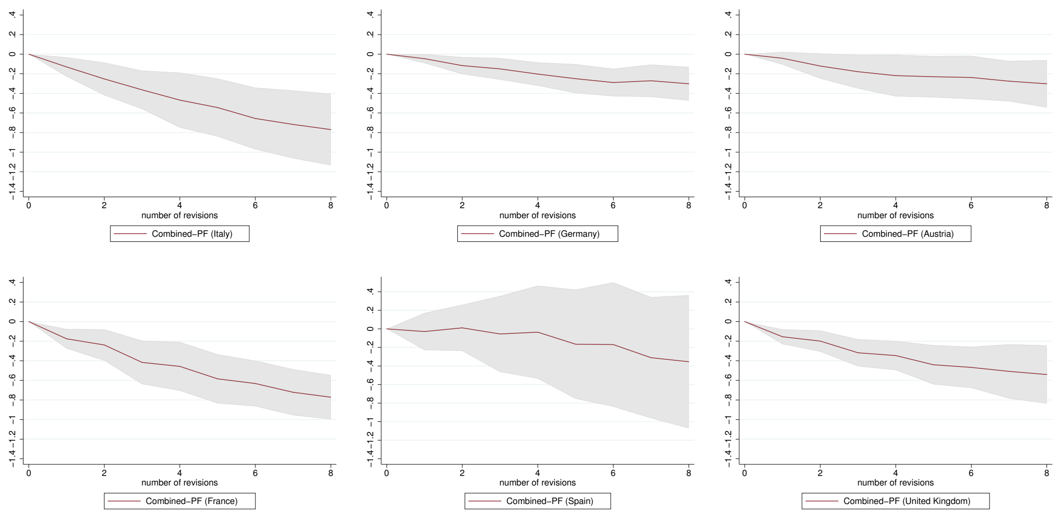
The estimation results for equation (1.4) are shown in Figure 1.2. The red line plots the estimates of α against the number of revisions $j - 1$. The shaded area around the red line is the 95% confidence interval. For all countries the estimated structural budget balances are systematically revised downward, implying that the initial ex post release is too optimistic. Hence, in $t + 1$ the fiscal position for year t looks better than a few years later. This overestimation is the result of an underestimation of the output gap and hence of the cyclical component of the budget balance. As the output gap for a certain year t is systematically larger in $t + j$ than in $t + 1$, the structural budget balance for year t is smaller in $t + j$ than in $t + 1$.

For Italy and France Figure 1.2 shows that the structural budget balance gets systematically revised downward by 0.8 percentage points of potential output after 8 revisions, i.e. after 4 years. In the United Kingdom, Germany, Austria and Spain the downward revision is somewhat less pronounced with 0.3 to 0.5 percentage points.[8] Apart from Spain all downward revision after 4 years are statistically

[8]Note that the cyclical component ($\epsilon_{B,GAP}\, GAP$) in our sample has an average absolute size of 0.8 percent of potential output, which is only twice the size of this revision size. In other words, the systematic size of these revisions is more than half of the absolute value of the cyclical component.

significant at a 5% level, implying that the method for estimating the structural budget balance is biased. For Italy, France and the United Kingdom this bias is already significant after the first revision, for Germany after four revisions and for Austria after five revisions. The large confidence bands for the Spanish bias reflects the huge revisions of the structural budget balance estimations for Spain, that have already been detected in the previous Section.

Figure 1.2: Bias of the ex post revisions of the structural budget (including 95% confidence interval) with respect to the initial estimate using the production function approach (in % of potential output)



Considering the importance of structural budget balances for the general fiscal surveillance and especially for the EU fiscal rules, even a small bias can be seen as highly problematic. Thus, a structural budget balance for year t that has been in line with the EU fiscal rules in $t+1$, might not be in line any more some time later in $t+j$ with $j > 1$. This finding of an overestimation of the structural budget balance and its large revisions is in line with the findings of Kempkes (2014), Marcellino and Musso (2011) and others, who focused their analyses on output gaps and showed that these estimations are typically revised upwards over time.

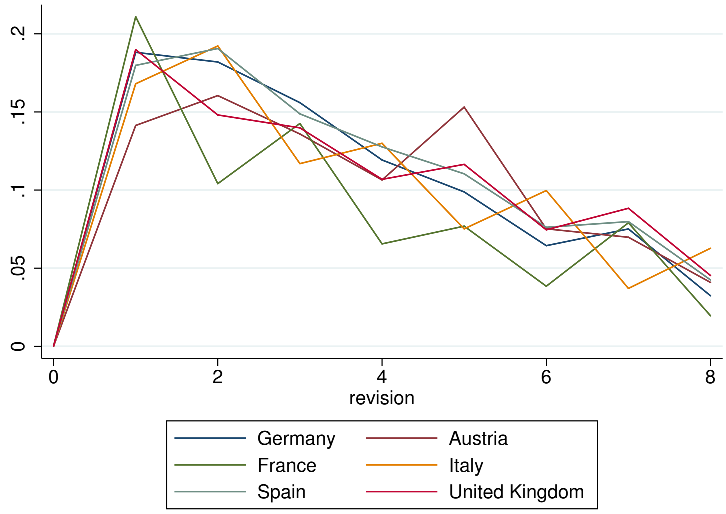
1.4 Revisions of the Structural Budget Balance: Time Series Approach

1.4.1 Data and Revision Size

To avoid the problems of the production function approach it is sometimes proposed to replace it with a time series approach (McMorrow *et al.*, 2015). However, we will show that this method suffers from the same shortcomings as the production function approach and that the resulting structural budget balances are biased and subject to large ex post revisions. While the EC publishes its own estimates of trend output from a Hodrick-Prescott (HP) filter, we constructed the OECD and IMF estimates by applying the HP filter with a smoothing parameter of $\lambda = 100$ on annual GDP series which are extended by the projections published in the OECD Economic Outlook and the IMF World Economic Outlook (see Maravall and del Río, 2001; ECB, 2007, for more details on the HP filter). The projections are included so as to mitigate the influence of the end-point problem (Kaiser and Maravall, 2001).

The structural budget balances are calculated as described in the previous Section. The revisions in absolute terms are displayed in Figure 1.3. The average size of the first revision for Germany, Austria and France are 0.19 percentage points of potential output, 0.14 and 0.21, respectively, and thus somewhat larger compared to the production function approach. For Italy and the UK the first revision still amounts to 0.17 and 0.19 percentage points of potential output. Only for Spain, the revision size is with 0.18 percentage points smaller than in the production function approach. While the revision size diminishes over time, it still remains sizeable even after four years ($j = 9$). In contrast to the production function approach, however, the revision sizes of the time series approach decline faster and lie between 0.02 and 0.06 percentage points of potential output after eight revisions.

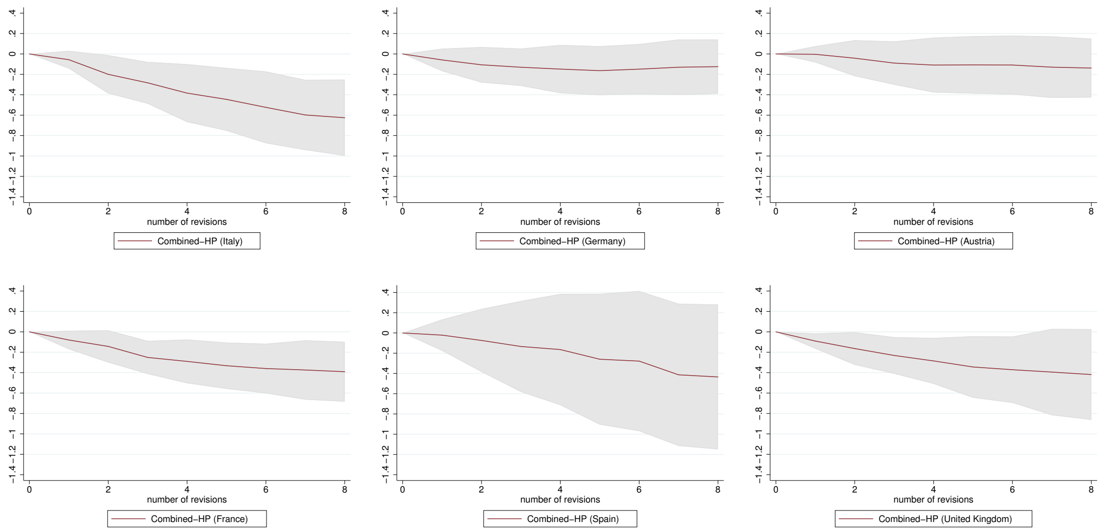
Figure 1.3: Size of the ex post revisions of the structural budget balance between two subsequent vintages using an HP filter (in percentage points)



1.4.2 Biasedness

The bias of the ex post revisions of the structural budget balance with respect to the initial estimate is shown in Figure 1.4. It was estimated using the same setting as in Section 1.3.2. For Italy, France and the United Kingdom, the bias is now smaller compared to the production function approach. In contrast to the production function approach, the downward revisions for Germany are not significant at a 5% level. Similar to the production function approach, there is no significant bias for Spain at a 5% level since the confidence bands are very large.

Figure 1.4: Bias of the ex post revisions of the structural budget balance with respect to the initial estimate using the an HP filter (in % of potential output)



1.5 Revisions of the Structural Budget Balance: Survey-Based Approach

The previous two Sections have shown that the mainstream methods of estimating structural budget balances are very problematic from a policy perspective as the resulting estimates are upward biased and subject to large ex post revisions. In this Section we present a novel survey-based approach that largely overcomes these problems.

1.5.1 Construction of the Survey-Based Structural Budget Balance

Empirical studies show that there is a high correlation between measures of the output gap and the degree of capacity utilization in the economy (Orphanides *et al.*, 2000). This is hardly surprising as according to the Bundesbank (2014, p. 14) "potential output is generally defined as that level of activity that occurs when capacity utilization in the economy as a whole is 'normal'." If capacity utilization in the economy is above normal, the level of activity as measured by gross domestic product is larger than potential output ($GDP > POT$) and the output gap is positive. The opposite holds for a negative output gap ($GDP < POT$), with capacity utilization being below normal. Given this high correlation it is obvious to propose an approach that uses the degree of capacity utilization to identify the cyclical component of the budget balance and to estimate the structural budget balance.

In many countries measures of capacity utilization are collected through business surveys. In the EU a representative sample of firms in services and manufacturing, which accounts for roughly 90% of gross value added, are asked for their capacity utilization on a quarterly basis.[9] According to the ECB (2007, p. 47, fn. 2), "survey respondents provide an answer about their overall resource utilisation, i.e. they consider both capital and labour inputs. This assumption is based on the overall content of the survey, which explicitly asks about various production constraints, including shortages of capital, labour and other inputs, suggesting that respondents have all those production inputs in mind when evaluating their capacity utilisation". The series for capacity utilization in manufacturing are available at least since 1996 and those for services at least since 2011 (see Table 1.1). Capacity utilization of the aggregate economy is the weighted average of both sectors.[10]

Since the degree of capacity utilization itself is not sufficient to identify an over- or underutilization of the economy, a normal level has to be determined. Here we follow Marcellino and Musso (2011) and assume that an economy's normal level of capacity utilization is given by the mean of the time series.[11] Thus, for every country i and year t the capacity utilization gap ($CUG_{i,t}$) is calculated as the deviation of the actual degree of capacity utilization ($CU_{i,t}$) from its mean ($\overline{CU_i}$):

$$CUG_{i,t} = CU_{i,t} - \overline{CU_i}. \tag{1.5}$$

If $CUG_{i,t}$ is positive, the economy's capacities are currently over-utilized, if it is negative, they are under-utilized.

The values for $CU_{i,t}$ are not revised after their initial publication. Thus, the only source of revision of $CUG_{i,t}$ is a change in $\overline{CU_i}$, which happens with every new observation. Obviously, a longer series with more observations has a more stable mean. As the series for capacity utilization in the service sector are relatively

[9] The question, which is asked in January, April, July and October, reads: "At what capacity is your company currently operating (as a percentage of full capacity)?" (ECB, 2007).

[10] We use the same shares as used for the construction of the Economic Sentiment Indicator (ESI), which is 4/7 for the manufacturing sector and 3/7 for the service sector.

[11] One could also assume that this normal level varies over time, i.e. by applying moving averages or filter methods to the degree of capacity utilization (ECB, 2007). However, in particular the latter would be subject to the already mentioned problems of the filtering methods. Nevertheless, we can assume that the revisions would be smaller than the ones of the mainstream approaches since there are no data revisions for capacity utilization. Silva *et al.* (2016) and Szörfi (2015) used survey information (i.a. capacity utilization) in an unobserved components model. Their measure had already a smaller revision size than the mainstream methods. However, a direct use of capacity utilization as proposed in Marcellino and Musso (2011) or suggested in Orphanides *et al.* (2000) further decreases the revision size substantially as we will demonstrate.

Table 1.1: First publication of capacity utilization data

Sector	IT	DE	AT	FR	ES	UK
Manufacturing	1985q1	1991q1	1996q1	1991q1	1987q2	1985q1
Services	2010q1	2011q2	2011q3	2011q4	2011q3	2011q2
Services (Backcasted)	1998q1	1995q2	1996q4	1988q1	1996q4	1997q1

Source: Eurostat.

short, we use other survey data from the service sector to backcast them as long as possible and by this to enlarge the series and to stabilize the mean (see Appendix A.2 for details). Figure 1.5 compares the latest estimate of the capacity utilization gap with the latest vintages (of spring 2018) of the output gaps estimated by the EC, the OECD and the IMF. For all countries, the capacity utilization gap is close to or within the range of these estimates represented by the shaded area. Thus, the production function-based output gaps and the capacity utilization gap are highly correlated. The correlation coefficients between the latest vintage of the capacity utilization gap and the output gap resulting from the production function approaches of the EC, the OECD and the IMF for the complete panel of countries ranges between 0.55 and 0.71 (see Table 1.2). The correlations with the output gaps of the EC are higher than with those of the OECD or the IMF. The correlations are substantially lower for the UK than for the other countries.

Table 1.2: Correlation between latest vintage for capacity utilization and latest vintages of latest vintage of traditional output gaps

	Panel	IT	DE	AT	FR	ES	UK
EC	0.709	0.733	0.831	0.737	0.739	0.758	0.563
OECD	0.616	0.641	0.773	0.725	0.775	0.626	0.477
IMF	0.545	0.571	0.732	0.598	0.745	0.558	0.316
Obs.	132	20	22	21	27	21	21

Figure 1.5: Capacity utilization gaps (red line) and range of latest vintages of output gaps (in % of potential output or percentage point deviation from the mean)

In order to be able to use the capacity utilization gap for the cyclical adjustment of the budget balances a value for the semi-elasticity of the budget balance with respect to the capacity utilization gap $\epsilon_{B,CUG}$ has to be calculated.

It is defined as the product of the semi-elasticity of the budget balance with respect to the output gap $\epsilon_{B,GAP}$ and the semi-elasticity of the output gap with respect to the capacity utilization gap $\epsilon_{GAP,CUG}$:

$$\epsilon_{B,CUG} = \epsilon_{B,GAP}\epsilon_{GAP,CUG}. \tag{1.6}$$

While for $\epsilon_{B,GAP}$ we use the values provided by the EC (see Section 1.3.1), the semi-elasticity of the output gap with respect to the capacity utilization gap is defined as

$$\epsilon_{GAP,CUG} = dln(GAP + 1)/dln(CUG + 1) \tag{1.7}$$

and estimated via

$$ln(GAP_{i,t}) = \beta ln(CUG_{i,t}) + u_{i,t}. \tag{1.8}$$

For every country i the results for equation (1.8) are summarized in Table 1.3. As already indicated by the high correlation between GAP and CUG in Figure 1.5, the estimates for β are significant at the 1% level for all countries. Also the R^2 are high enough to justify the construction of the semi-elasticity this way, especially

Table 1.3: Results for equation (1.8)

	IT	DE	AT	FR	ES	UK
CUG	0.833***	0.582***	0.513***	0.735***	1.078***	0.583***
	(0.169)	(0.108)	(0.107)	(0.138)	(0.202)	(0.190)
Obs.	20	22	21	27	21	21
Adj. R^2	0.537	0.560	0.513	0.502	0.567	0.286

Notes: Standard errors in parentheses. Level of significance: *** p<0.01, ** p<0.05, * p<0.1.

when comparing it to the goodness-of-fit of the estimations of the semi-elasticities in the production function approach (e.g. as in Price *et al.*, 2014). The estimates for β are then plugged into equation (1.6) as values for $\epsilon_{GAP,CUG}$. Finally, the survey-based structural budget balance is computed by adjusting equations (1.1) and (1.2) to

$$X = B - \epsilon_{B,CUG} CUG - T. \tag{1.9}$$

Also the survey-based structural budget balance is expressed in % of potential output. By using this most common scaling we simplify the comparability of the methods and the implementation of the new approach.[12]

Figure 1.6 compares the survey-based structural budget balance with those resulting from the production function approach using the vintages of spring 2018 for each method. For all approaches we subtracted the one-offs published by the EC, which are available since 2010. This enables us to focus our analysis on the differences that result from the estimation of the cyclical components. Prior to 2010 we only display the cyclically-adjusted balance, which is identical to the structural budget balance if one-offs are ignored. For most of the period the results of both approaches are quite similar. However, at the end of the sample the survey-based structural budget balance is smaller for almost all countries. Thus, for a given budget balance the fiscal space turns out to be tighter in the survey-based approach than in the production function approaches. However, as we already know from Section 1.3 the initial estimates of the structural budget balances resulting from the production function approaches are systematically revised downward in subsequent estimates. Thus, in a few years it is very likely that the discrepancy at today's end of the sample will disappear, provided the initial estimates of the survey-based approach are at least less biased.

[12] We could also use actual output instead of potential output. Both scaling methods lead to similar structural budget balances as long as the output gap is not too large (Fedelino *et al.*, 2009).

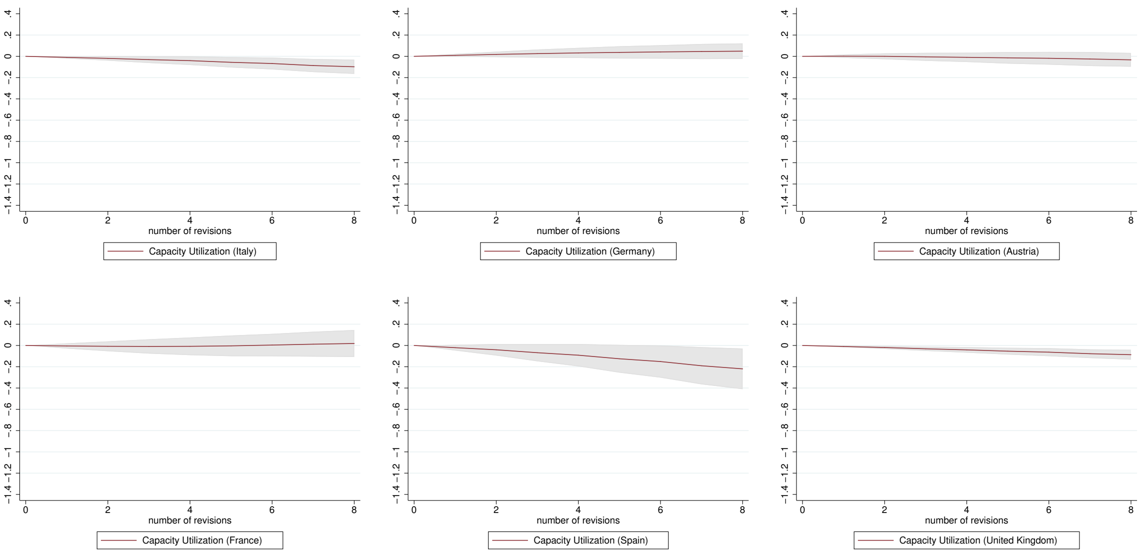
Figure 1.6: Structural and cyclically-adjusted budget balance: Survey-based (continuous and dashed red line) versus range of latest vintages from production function approach (in bn. euros)

1.5.2 Evaluation of the Survey-Based Approach

For estimating the bias of the survey-based structural budget balance, we re-estimate equation (1.4). The results are shown in Figure 1.7. The red line plots the estimates of α against the number of revisions $j-1$. The shaded area around the red line is the 95% confidence interval. For Germany, Austria and France the estimated structural budget balances are unbiased at the 5% level over all revision horizons. For Italy and the United Kingdom the estimates are unbiased up to the third revision. Then, the structural budget balance is systematically revised downward, which after 8 revisions amounts to 0.1 percentage points compared to the initial estimate. The results for Spain show a similar picture. Only the size of the downward revision is with 0.2 percentage points larger. A comparison with the estimated biases in both the production function approach (Section 1.3) and the time series approach (Section 1.4) however reveals that those biases in the survey-based approach which are statistically significant are quantitatively much smaller.[13]

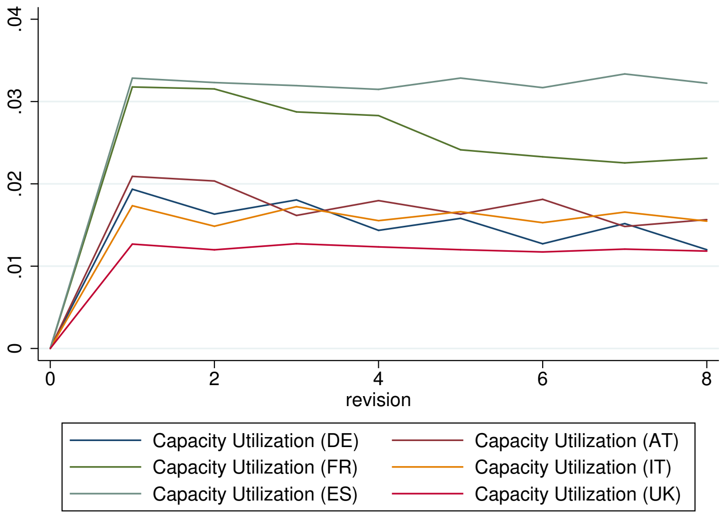
[13]The bias of the production function approach is on average over 8 horizons and all production function approaches by all institutions 10.6 times the size of the survey-based approach for Italy, 9.2 times for the United Kingdom and 1.0 times for Spain. The bias of the time series approach is on average 7.9 times the size of the survey-based approach for Italy, 6.8 times for the United Kingdom and 1.8 times for Spain.

Figure 1.7: Bias of the ex post revisions of the structural budget with respect to the initial estimate using the survey-based approach (in % of potential output)



The tight confidence intervals in Figure 1.7 already hint at a small revision size of the survey-based structural budget balance. This result is supported by the direct calculation of the size of the ex post revisions as shown in Figure 1.8. The revision size is 0.02 percentage points of potential output for Austria, Germany and Italy and 0.1 percentage points of potential output for the UK at the first revision. Even in the case of Spain and France, where the revision size turns out to be largest, the survey-based approach performs better than both the production function approach and the time series approach. It declined from 0.28 percentage points (production function approach) and 0.18 percentage points (time series approach) to 0.03 percentage points for the first revision for Spain. For France, it declined from 0.20 percentage points (production function approach) and 0.21 percentage points (time series approach) to 0.03 percentage points for the first revision. This small revision size in the case of Spain in the survey-based approach comes at the cost of having a statistically significant bias at a 5% level. However, we believe that this cost is acceptable since the large revision size in the mainstream approaches for Spain made them hardly usable at all. In addition, this bias will vanish when the capacity utilization series will be long enough. In the case of the United Kingdom and Italy this will happen even sooner. Furthermore, it has to be emphasized, that this bias is merely a statistical one. Quantitatively, it is negligible.[14]

Figure 1.8: Size of the ex post revisions of the structural budget balance between two subsequent vintages using the survey-based approach (in percentage points of potential output)



1.6 Conclusions

The budget dispute between Italy and the European Commission in 2018 gave new impetus for the debate about the reliability of output gap estimation methods and their use for calculating structural budget balances. In this paper we review the main properties of the mainstream approaches. We show that the structural budget balances resulting from the production function approach and the time series approach are imprecise, subject to large revisions and often biased. But apart from these technical flaws the mainstream approaches also suffer from political economy problems. As the computation of structural budget balances in the mainstream approach is difficult and model-dependent, it is not easy to explain to the public and prone to manipulation. In addition, the first ex post estimation is only available late with the first publication of GDP.

We therefore propose an alternative approach to calculate structural budget balances on the basis of a business survey. We show that compared to the main-

[14]An alternative interpretation of Figure 1.7 is the decline of capacity utilization during the Great Recession. That lead to a successive decline in $\overline{CU_i}$ in real time due to the small sample length.

stream approaches the structural budget balances are significantly less biased or even unbiased, more precise and only subject to very small revisions. They can be easily computed and explained to policy makers and the public. As the cyclical position is determined through a survey among firms, its outcome is hardly manipulable. Finally, the capacity utilization of a year is already available at the end of that year. So, the computation of the ex post structural budget balance only depends on the publication date of the unadjusted budget balance.

The survey-based structural budget balances can be used to improve fiscal surveillance in the EU in general, but in particular the evaluation of compliance with EU fiscal rules. Currently, the assessment of compliance with those goals is based on an upward-biased estimation as shown in this paper. The survey-based approach would provide an unbiased (or at least much less biased) measuring tool for the concept of structural budget balances and their ex post evaluation. But also for the budgetary goals, which are formulated ex ante our method can be of use. At the end of the first month of a quarter, the capacity utilization of the respective quarter is available. The missing quarters can be estimated via standard forecasting methods already in use in various institutions. This would further help to improve the current evaluation of the fiscal rules since the estimation of output gaps (and consequently the estimation of structural budget balances) is also biased when assessing ex ante formulated output gaps (Kempkes, 2014).[15] We do not want to provide a forecasting method at this point. This is left for future research. Moreover, some parts of the preventive arm depend on the current position in the business cycle that is assessed by the output gap. Here, the mainstream output gap could be replaced by the survey-based capacity utilization gap.

The methods of estimating the structural budget balance have been often revised in the EU in the past. In the knowledge of their shortcomings (e.g. their biasedness and low precision), more difficult methods that are still biased and imprecise have been introduced. That lead to a decrease of the legitimacy of structural budget balance as a part of the fiscal framework and opened the door for proposing estimation methods that systematically give a member state more fiscal space. The survey-based approach can help to strengthen the legitimacy of the existing rules by overcoming the shortcomings of the mainstream approaches and thereby suppressing politically motivated proposals of new approaches that merely have the purpose of undermining the SGP.

[15] Although, we only evaluated the use of ex post estimated structural budget balances, it can be expected that our approach would be also unbiased (or at least less biased) for ex ante estimated structural budget balances. The main difficulty in the mainstream approaches is the split in a structural and a trend component. Our approach is not affected by this problem.

Chapter 2

Tax Revenue Forecast Errors: Wrong Predictions of the Tax Base or the Elasticity? [1]

2.1 Introduction

Tax revenue forecasts are an essential input for fiscal policy planning in the short- and medium-term (Auerbach, 1999), especially if public spending is restricted by special debt brakes or fiscal rules in general. Given that tax revenue forecasts are not systematically biased in any direction, they can serve as a guideline for future spending leeway and prevent governments to run into excessive public debts and deficits. Despite political economy arguments, biases in tax revenue estimates largely stem from wrong macroeconomic predictions. If the macroeconomic outlook changes or was assessed wrongly, this ultimately translates into false tax revenue predictions. However, not only the prediction of macroeconomic conditions itself leads to biased estimates on future tax revenues, but also wrong assessments of linkages across macroeconomic aggregates and tax revenues, namely, tax revenue elasticities. In this paper, we disentangle tax revenue forecast errors into these two sources of possible distortions for a variety of tax types and the overall tax sum.

The accuracy of overall tax revenue estimates has often been investigated for a wide range of countries in different setups (see, for a survey, Leal *et al.*, 2008). However, the consequences for the forecast accuracy of the interplay between macroeconomic assumptions and their transmission into tax revenue predictions is missing. Based on his own findings, Heinemann (2006) hypothesized that wrong budgetary

[1] This chapter is joint work with Robert Lehmann. An older version circulates as Göttert and Lehmann (2021).

forecasts are mainly driven by false macroeconomic growth projections and wrong assumptions on the tax elasticities. He, however, does not go deeper in this direction by disentangling both sources of errors. We take up this road and study tax revenue estimates at the federal level of Germany as those have a long tradition and receive a considerable media attention. Ademmer and Boysen-Hogrefe (2022) also contribute an interesting finding: tax revenue forecast errors in the medium-term considerably translate into the budget balance of German states, thus, overoptimistic forecasts drive total debt. Furthermore, Germany gives us a very interesting setup as tax revenue estimates are produced by an official board or working group that covers governmental agents as well as independent experts. Single tax revenue predictions are produced by each independent expert in advance and the final numbers are reached via a consensus across the working group's members. The macroeconomic outlook, however, is given by the governmental agents and is thus exogeneous to the working group. So in the end, the resulting forecast error of the working group is driven by the accuracy of the exogeneous macroeconomic assumptions and the translation of these figures into final tax estimates.

Our results indeed suggest that both error sources matter, with heterogeneous degrees across the tax types. For profit-related taxes as well as the wage tax, more than 90% of the explained forecast error can be attributed to wrong macroeconomic assumptions. The opposite holds true for the energy tax and sales taxes. 94% of the explained energy tax forecast error and two-thirds of the explained sales taxes error is attributed to a false assessment of the tax revenue elasticity with respect to its corresponding tax base. For the overall tax sum, 31% of the explained error can be attributed to the forecast error of the elasticity and 69% to the error of a wrong macroeconomic prediction.

Given our findings, we complement the existing literature for Germany in two ways. First, we investigate the channel of a rather methodological influence and formulate a statement on its absolute and relative importance compared to forecast errors stemming from macroeconomic projections. Second, we not only analyze the overall tax sum but rather broaden the picture and investigate the forecasting performance for the six largest taxes in Germany that account, on average, for 81% of total tax revenues in the last three decades. An early and in depth analysis on the accuracy of tax revenue forecasts in Germany is provided by Gebhardt (2001). For the period from 1970 to 2000, he attests the working group a high average short-term forecast accuracy (here: the running year) but cannot fully disentangle the forecast errors due to data restrictions. Büttner and Kauder (2015) find similar results for total tax revenues and the period running from 1971 to 2013. The forecasts for the current and the following year are, on average, unbiased, but forecast errors for gross domestic product (GDP) explain lots of the variation in total tax revenue forecast errors. The studies by Heinemann (2006) and

Breuer (2015) put the spotlight on the medium-term performance of tax revenue estimates. Heinemann (2006) summarizes that revenue predictions are biased towards overoptimism, especially for forecast horizons longer than one year ahead. Breuer (2015) underpins the finding that medium-term tax projections are too optimistic, which is mainly described by forecast errors for GDP. Bischoff and Gohout (2010) find that total tax revenues for the states belonging to former Western Germany are generally not upward biased, but the overoptimism is negatively correlated with the incumbents' popularity and thus their chances of being reelected. Finally, Kauder *et al.* (2017) study the impact of state elections on the accuracy of regional fiscal forecasts. They find that tax revenue forecasts are underestimated by Eastern German state governments in years prior to an election.

German tax revenue predictions have also been part of studies that exploit the international dimension of forecast accuracy and its determinants. A large part of the existing studies deal with political economy arguments (see, among others, Beetsma *et al.*, 2009, 2013; Cimadomo, 2016; Jochimsen and Lehmann, 2017; Strauch *et al.*, 2004). Büttner and Kauder (2010) instead explore how cross-country differences in the tax revenue forecasting process (for example, the length of the forecast horizon, the number of taxes predicted, and, more important, the overall independence of the preparation) correlate with the degree in forecast accuracy. Despite the large influence of wrong macroeconomic projections, they find that the degree in independence is positively associated with the forecast accuracy. It seems therefore reasonable to outsource the whole forecasting process to independent institutions, which has also been emphasized by Jonung and Larch (2006). We therefore compare the accuracy of macroeconomic projections from the working group with those of an independent institution in Germany that serve as input for the government. It turns out that the independent forecaster outperforms the predictions of the government, thus, the tax revenue forecast errors can, c.p., be reduced by fully outsourcing the macroeconomic projections to the already existing independent institution.

Our paper is structured as follows. In Section 2.2, we introduce the institutional framework in which German tax revenue forecasts operate. Section 2.3 outlines the theoretical considerations guiding our empirical strategy (Section **??**). In Section 2.5, we present our main results, followed by the comparison of accuracy to the independent forecasting institution in Section 2.6. The last section concludes.

2.2 Institutional Setup

Since 1955, tax revenue forecasts in Germany are produced by the Working Party on Tax Revenue Estimates (WPTRE, in German: *Arbeitskreis Steuerschätzungen*). The Working Party acts as an advisory board for the Federal Ministry of Finance and grounded its acting on bylaws in the mid of 2017.[2] To ensure the independence, the Party consists of 27 governmental and non-governmental members: the Federal Ministry of Finance (head), the Federal Ministry for Economic Affairs and Energy, the five main Economic Research Institutes in Germany[3], the Federal Statistical Office, the German Central Bank, the German Council of Economic Experts, each of the 16 Federal Ministries of Finance, and the Association of German Cities.

The Working Party meets twice a year, namely in May and in November. In May, the forecasts are formulated for the current and the four upcoming years and serve as the basis for the budget draft of the following year and the annual update of the medium-term budget planning. The forecast in November comprises the current and the five upcoming years and serves as the final formulation of next year's tax revenues in the federal budget. The Working Party follows a bottom-up approach, meaning that each of the 32 single taxes (resp. tax groups)[4] is forecasted separately. At each forecasting date, eight members—the five Research Institutes, the German Central Bank, the German Council of Economic Experts, and the Federal Ministry of Finance—present their individual forecasts and a discussion takes place for each single tax (resp. tax groups) until a consensus is reached among all members.

Each individual forecast is calculated under two circumstances. Whereas the Federal Ministry of Finance allows each member to base their forecasts on their individual methodology, the macroeconomic projections (for example, the growth of nominal gross domestic product) are given by the political authority; this applies equally to all tax bases. Thus, each member has to base its tax revenue forecasts on the macroeconomic projections of the government. This is the main reason to let Büttner and Kauder (2010) state that German tax revenue forecasts are by no means fully independent. In general, the individual tax revenues are calculated according to the existing taxation law. However, planned tax law changes are estimated by the Federal Ministry of Finance and are included in the budgetary planning.

Not much information is available about the applied methodology of each Party member. Büttner and Kauder (2008) published a book that compares tax revenue forecast methodologies for a multitude of countries. According to their descrip-

[2] A detailed description of the Working Party on Tax Revenue Estimates can be found here: `https://www.bundesfinanzministerium.de/Content/EN/Standardartikel/Topics/Taxation/Articles/working-party-on-tax-revenue-estimates.html` (accessed on August 12, 2020).

[3] DIW: German Institute for Economic Research in Berlin; IfW: Kiel Institute for the World Economy; ifo: ifo Institute–Leibniz Institute for Economic Research at the University of Munich e.V.; RWI: RWI Leibniz Institute for Economic Research; IWH: Leibniz Institute for Economic Research Halle.

[4] A list containing the 32 taxes (resp. tax groups) can be found in Appendix B.1.

tions, the members of the Working Party on Tax Revenue Estimates mainly base their forecasts on *indirect* methods such as macroeconomic simulations or the application of tax elasticities. Macroeconomic simulations make usage of the existing tax law, tariffs and their connection to various tax bases. In contrast to microeconomic simulations, all functional forms are specified at the macroeconomic instead of the individual level. Elasticity methods are a common practice: based on historical data, an elasticity between a tax type and its corresponding tax base is estimated and applied for forecasting (Büttner and Kauder, 2008, 2010). Both methods can coincide, especially if the elasticity model is specified correctly. As the macroeconomic simulation postulates functional forms between a tax type and a tax base—given the current law and existing tax allowances—it implicitly models elasticities.

In the end, a consensus is reached among the party members. After the completion of a forecast meeting, the results are published to inform the interested public. The federal government uses the forecasts for its budget draft as well as its medium-term financial planning. The results are also the basis for the estimation of state-specific tax revenues and serve as a major indicator for many municipalities.

2.3 Theoretical Considerations

The Working Party on Tax Revenue Estimates in Germany typically focuses on a multitude of target series to forecast, ranging from $R = 32$ different taxes (resp. tax groups), y_t^r, to total tax revenues, $Y_t = \sum_{r=1}^{R} y_t^r$. Prior to the first release of the tax revenue figures for a specific year t, the WPTRE produces its forecasts, $\widehat{y}_{t|t-h}^r$ and $\widehat{Y}_{t|t-h}$, at different points in time with a forecast horizon of h years. After the release, the WPTRE's forecast errors, $\mathrm{FE}_{h,t}^r = \widehat{y}_{t|t-h}^r - y_t^r$ and $\mathrm{FE}_{h,t} = \widehat{Y}_{t|t-h} - Y_t$, can be calculated. A positive sign indicates an overestimation; the opposite represents an underestimation.

Assume that the Working Party is fully rational and tries to formulate an optimal forecast at $t-h$, given an information set Ω_{t-h}; this set comprises information on the different tax revenues such as the corresponding tax bases. If the WPTRE operates under a quadratic loss function, it tries to minimize the expected mean squared forecast error (Batchelor, 2007)

$$\mathcal{L}_h^r = \mathrm{E}\left[\left(\widehat{y}_{t|t-h}^r - y_t^r\right)^2 |\Omega_{t-h}\right] . \tag{2.1}$$

The optimal forecast, $\widehat{y}_{t|t-h}^{*,r}$, is than given as the conditional expectation (Nordhaus, 1987)

$$\widehat{y}^{*,r}_{t|t-h} = \mathrm{E}\left[y^r_t|\Omega_{t-h}\right] \ . \tag{2.2}$$

This forecast, and thus the resulting forecast error, mainly depends on the information set. At time $t-h$, the WPTRE produces its forecast given that the prediction is a function of the information set: $\widehat{y}^r_{t|t-h} = f(\Omega_{t-h})$. According to the institutional setup described in Section 2.2, we can model the information set in much more detail. We distinguish between both exogenous and endogenous factors that the WPTRE faces by producing its forecasts.

The applied methodology (METHOD) of each Party Member can be treated as an endogenous factor. If research activities take place, each Party Member might adopt new methodologies over time. However, this endogenous factor might be superimposed by one exogenous factor: the final forecast of the Working Party is a consensus (CONS) among all members, thus, the member-specific forecast is weighted by its own bargaining power within the Party. As we have no further insights on the distribution of power within the Working Party, we assume the weights to be equally distributed. Also the composition (COMP) of the Working Party is exogenous and time-invariant for the time of our analysis. Since the basic estimations of the working party members are on the basis of the current tax law, planned tax law changes (LAW) need to be taken into account additionally. They are provided by the Federal Ministry of Finance and are exogenous for the party members.[5] The last exogenous factor is crucial for our analysis, namely the Federal Ministry's input of the tax base forecasts (BASE).

All these factors are part of the information set that the Working Party faces at each forecasting date. It reads as: $\Omega_{t-h} = \{$ METHOD, CONS, COMP, LAW, BASE$\}$. The final forecast of the WPTRE therefore is a function of these factors and, thus, the forecast error is: $\mathrm{FE}^r_{h,t|\Omega_{t-h}} = \widehat{y}^r_{t|t-h,\Omega_{t-h}} - y^r_t$. In the following, we assume—as the tax base forecasts are given to the Working Party—that the published forecasts are produced by a member-weighted elasticity method; the consensus among the members leads to a final forecast that can be linked to the tax base input provided by the Federal Ministry for Economic Affairs and Energy. We also argue that this assumption is realistic from a forecaster's point of view that works in the field of public economics. As elasticities can be seen as central figures in this field, the forecaster might run a cross-check by calculating the elasticities after the production of the forecasts. If these resulting elasticities seem implausible and heavily deviate from any given anchor such as long-term averages, an adjustment of the forecasts might take place before the final publication. As

[5]The estimation of planned tax law changes and the general tax estimation are executed by different sections of the Federal Ministry of Finance. Effects of planned tax law changes on macroeconomic figures are taken into account in the macroeconomic projection by the Federal Ministry for Economic Affairs and Energy.

already stated in the previous section, elasticity methods are a common practice among the party members. Thus, the final outcome of the discussion within the party can be treated as an equally-weighted forecast based on member-specific elasticity estimates.

We use the standard tax revenue elasticity, $\varepsilon^r_{y,B}$, which links the development of tax r to the development of its specific tax base B. This elasticity is defined as the ratio between the growth of tax revenues and the growth of the corresponding tax base:

$$\varepsilon^r_{y,B} = \frac{dy^r}{dB} \cdot \frac{B}{y^r} \,. \tag{2.3}$$

We can easily calculate these ex-post elasticities based on the given tax and macroeconomic data. If we, instead, take an ex-ante stand, we can replace each component by its forecast:

$$\widehat{\varepsilon}^r_{y,B} = \frac{\widehat{y}^r}{\widehat{B}^{\text{ex},r}} \,. \tag{2.4}$$

As previously stated, the forecast for the tax base is given to the Working Party and is thus exogenous ($\widehat{B}^{\text{ex}}$) as the Working Party has no veto against this prediction. The revenue forecast for each tax is reached via a consensus among the members, thus, we can calculate the resulting and implicit elasticity based on the published forecasts. It is, however, obvious that the forecast error of the Working Party mainly depends on the forecast error of the tax base as well as a wrong specification of the elasticity, given the other factors influencing the outcome. In the following section, we elaborate more on how we implement these theoretical considerations empirically.

2.4 Empirical Strategy

2.4.1 Derivation of the Empirical Model

As stated in Section 2.3, the forecast of the Working Party and thus the forecast error for a specific horizon h depends on the underlying information set, $\text{FE}^r_{h,t|\Omega_{t-h}} = \widehat{y}^r_{t|t-h,\Omega_{t-h}} - y^r_t$. This expression in growth rates leads to the first, rather standard check on unbiasedness in the sense of Holden and Peel (1990b), $\text{FE}^r_{h,t|\Omega_{t-h}} = c^r_h + \eta^r_{h,t}$, that is, the forecast error should not show any systematic positive or negative bias. If this is true, the horizon-specific constant, c^r_h, becomes zero; the idiosyncratic disturbances are denoted by $\eta^r_{h,t}$. However, unbiasedness is not the only criterion a forecast should fulfill. Weak rationality is the second characteristic which is discussed in the literature. The idea behind this characteristic is to ask whether the forecaster incorporates all possible information available at the time the forecast is produced. We, therefore, rearrange the forecast error and transfer it to an empirical application in the sense of Mincer and Zarnowitz (1969): $y_t^r = -c_h^r + \alpha \widehat{y}_{t|t-h,\Omega_{t-h}}^r - \eta_{h,t}^r$. Now we substract the forecast from both sides of the equation, rearrange the equation and get: $\text{FE}_{h,t|\Omega_{t-h}}^r = c_h^r - (1-\alpha)\widehat{y}_{t|t-h,\Omega_{t-h}}^r + \eta_{h,t}^r = c_h^r + \beta \widehat{y}_{t|t-h,\Omega_{t-h}}^r + \eta_{h,t}^r$; this empirical setup is constructed in similar vein to Keane and Runkle (1989, 1990) who investigate the forecast accuracy of price forecasts. If no systematic bias exists and the forecast error is not correlated with the forecast ($c_h^r = \beta = 0$), then the forecast fulfills the criteria of weak rationality. In this case, only the forecast itself plays a role in explaining the forecast error. There are, however, more information available at the time when the forecast is calculated. If all information are treated efficiently, then the forecast fulfills the criterion of strong rationality.

In our notation from Section 2.3, the forecast is a function of the available information set: $\widehat{y}_{t|t-h,\Omega_{t-h}}^r = f(\Omega_{t-h})$. Taking the previous argumentation, strong rationality reads as:

$$\text{FE}_{h,t|\Omega_{t-h}}^r = c_h^r + f(\Omega_{t-h}) + \eta_{h,t}^r \, .$$

For simplicity, we assume the information set function to have a linear form, depending on the factors introduced in the previous section: f(METHOD, CONS, COMP, LAW, BASE) with each factor modelled by a function itself. As argued, each member of the Party can have its own methodological approach, that is also time-dependent and could change over the years. However, this individual approach is outweighted as the final forecast is reached via a consensus across the members. In fact, the methodology is time-invariant, with the exception that an implicit elasticity ($\widehat{\varepsilon}_{y,B}^r$) results from the final forecast. Thus, the tax revenue forecast error might depend on a wrong assessment of the elasticity. We therefore assume that the influence of the methodology can be expressed by the forecast error of the tax revenue elasticity: $\text{METHOD} = \text{FE}_{h,t}^{\varepsilon}$.

The influence of reaching a consensus per tax is time-invariant and reads as: $\text{CONS} = \text{CONS}_h^r$. The same applies to the composition of the Working Party. It might be the case that all or a group of forecaster produce systematically higher or lower forecast errors for a specific tax, depending on the characteristics of the forecaster (for example, the priority of putting effort in developing better forecasting techniques for a specific tax). Thus, the influence of the composition on tax revenue forecast errors of the Working Party is time-invariant: $\text{COMP} = \text{COMP}_h^r$. The tax base forecasts, instead, are exogenous to the Working Party. As the party members have to take these forecasts from the Federal Ministry of Finance, wrong assessments on the development of the tax base are converted one-to-one into the tax revenue forecast errors. It might therefore be reasonable to think that the influence of the tax bases can be expressed as their forecast errors: BASE $= \mathrm{FE}^{B}_{h,t}$. Taking all these influences together, our empirical model reads as:

$$\mathrm{FE}^{r}_{h,t} = \beta_1 \mathrm{FE}^{B}_{h,t} + \beta_2 \mathrm{FE}^{\varepsilon}_{h,t} + a^{r}_{h} + u^{r}_{h,t} \ , \tag{2.5}$$

with a^{r}_{h} as the time-invariant characteristics of the forecast calculation and $u^{r}_{h,t}$ as the usual idiosyncratic error term. This is a standard panel fixed-effects representation and each of the models' parameters—for simplicity we apply the same notation for each model here—are estimated by standard OLS. The parameters' standard errors are robust and clustered at the horizon level.

2.4.2 Data Set

We examine forecasts in growth rates of the overall tax sum and the six largest taxes (resp. tax groups): wage tax, assessed income tax, business tax, corporate tax, energy tax and sales taxes.[6] Between 1991 and 2019 these six taxes contributed, on average, by 81% to the overall tax sum. All tax revenues are measured at the federal level.[7] We analyze the time-span between 1992 and 2019.[8] However, we have to restrict that range whenever the data availability is limited. For that reason we focus on the first four forecasting horizons. The WPTRE meets twice a year: usually in spring and in fall.[9] We analyze the forecasts for the current and for the following year. In spring, the latest realized macroeconomic values are the ones for the fourth quarter of the previous year. In fall, the latest published macroeconomic values are on the second quarter of the ongoing year. Consequently, we use the notation of horizon $h = 0.5$ for a forecast for the current year t produced in fall of the current year and $h = 1.5$ for a forecast for the next year $t + 1$. Similarly, the forecast produced for the current year in spring of the current year corresponds to horizon $h = 1$ and $h = 2$ stands for a forecast for the next year.

There are missing values for the corporate tax estimations for the years 2002 and 2003 at the first two horizons and for the year 2003 at the third horizon. Regarding the realized values, there is a severe outlier for the corporate tax with a decline of more than 700% in the year 2002. We excluded that year from our analysis of the corporate tax.

Figure 2.1 shows the forecasted (dashed line) and realized growth rates (solid line) of our sample beginning with the first horizon ($t + 0.5$) and ending with

[6] Appendix B.1 contains the data sources and the time spans for which the data are available.

[7] For discussions on the German states, see Kauder *et al.* (2017) and Bischoff and Gohout (2010).

[8] We restrict our sample to avoid structural breaks possibly driven by the German reunification.

[9] In 2020, there was an additional third meeting during fall due to the Corona crisis.

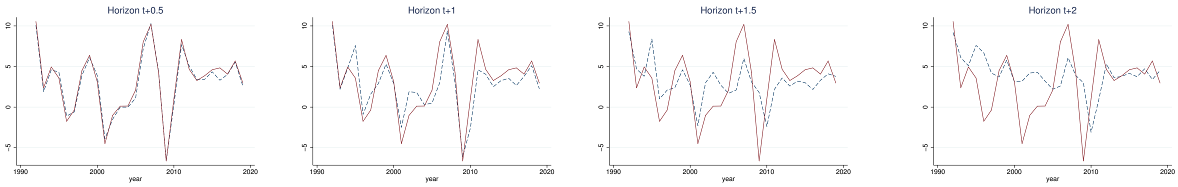
the last horizon ($t+2$). Already at first glance, it is obvious that the shorter the forecast horizon is, the smaller is the forecast error. One can neither spot any clear bias nor a tax that performs particularly bad except of the one very pronounced overestimation at the third horizon for the corporate tax. However, Section 2.5 will deal with these questions from an empirical stance. Furthermore, it deals with the question if we can assign the forecast errors, the differences between the dashed and the solid lines, to a wrong elasticity or a wrong tax base forecast.

The forecast errors for the underlying tax bases are calculated as the difference between the forecasts and their published values. The implicitly forecasted and realized elasticities of the taxes are calculated according to Equation (4). We assign the taxes to their corresponding tax base, which are usually applied by the WPTRE in its forecasting procedure; Gebhardt (2001) provides an overview of the applied tax bases by the WPTRE. Nominal GDP is the assigned tax base for the overall tax sum, gross wages and salaries for the wage tax, corporate and investment income for the assessed income, the business and the corporate tax, nominal GDP for the energy tax and private consumption for the sales taxes.[10] Although these are not the actual tax bases, the WPTRE uses the aforementioned macroeconomic variables as tax bases for its forecast due to data restrictions (Gebhardt, 2001).

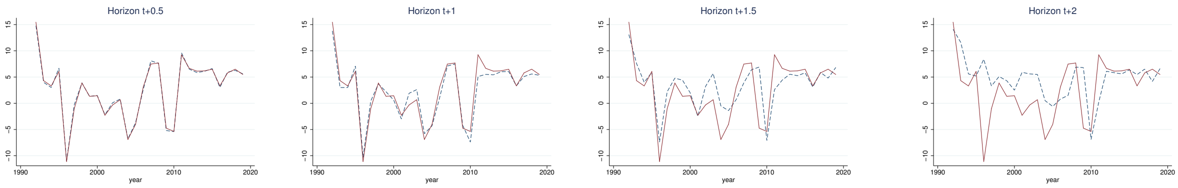
[10]There are two estimates of zero for the forecasted growth rate of the wages at the first horizon (2003 and 2005). Thus, the forecasted elasticities cannot be calculated at that horizon and are treated as missing.For the energy tax, Gebhardt (2001) did not report a tax base. We assigned it to nominal GDP since oil consumption and economic growth are closely correlated. The tax base reported for the sales taxes in Gebhardt (2001) is outdated. Consequently, we updated it.

Figure 2.1: Forecast performance for different taxes and horizons

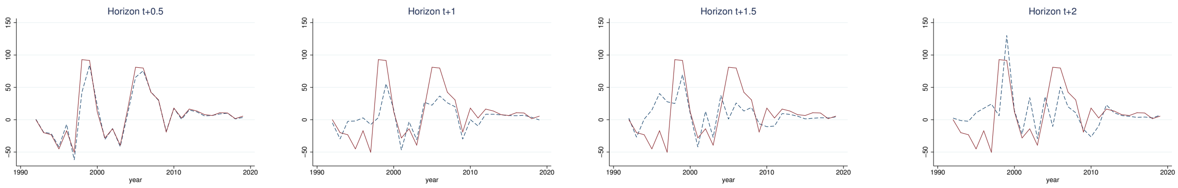
(a) Tax Sum



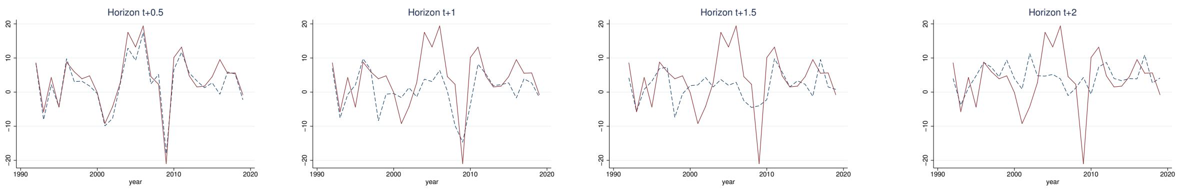
(b) Wage Tax



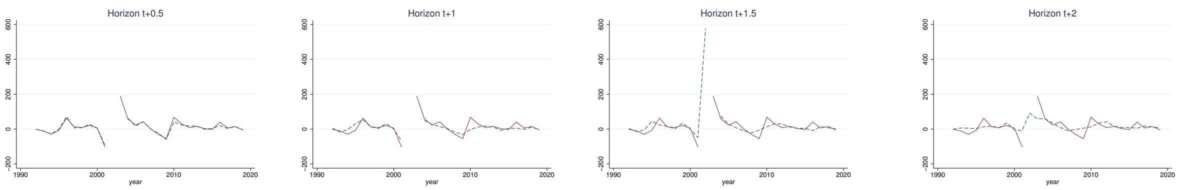
(c) Income Tax



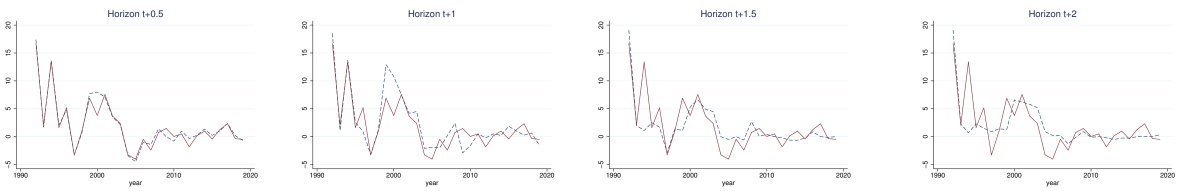
(d) Business Tax



(e) Corporate Tax



(f) Energy Tax



(g) Sales Taxes

Notes: The figures compare the forecasts (dashed line) with the realizations (solid line) of the growth rates (in %).

2.5 Results

2.5.1 Unbiasedness

To test for unbiasedness, we follow Section 2.4 and estimate the following panel model,

$$\mathrm{FE}^i_{h,t} = \alpha^i_1 + u^i_{1,h,t} \,, \tag{2.6}$$

with i either representing tax revenues ($i = r$), the tax bases ($i = B$), or the elasticities ($i = \varepsilon$). Table 2.1 shows the results for the tax revenue forecast errors. The overall tax sum is not systematically biased in any direction, which is in line with Breuer (2015) and Büttner and Kauder (2015). The results are mixed for the individual taxes. Whereas we cannot detect any bias for the wage, corporate and energy tax, the income and business tax are underestimated; the sales taxes forecasts are on average too optimistic.

Table 2.1: Test on Unbiasedness of Tax Revenue Forecasts

	Tax Sum	Wage Tax	Income Tax	Business Tax	Corpor. Tax	Energy Tax	Sales Taxes
α^r_1	0.05	0.64	-2.98*	-1.55*	0.37	0.11	0.44***
	(0.30)	(0.53)	(1.56)	(0.82)	(0.82)	(0.17)	(0.17)
Obs.	112	112	112	112	105	112	112
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Moreover, all key macroeconomic forecasts except corporate income show a bias (see Table 2.2). First of all, nominal GDP growth is overestimated by 0.41 percentage points in the sample period which is in line with the finding of Breuer (2015). Furthermore, the sum of gross wages and salaries as well as private consumption, which serve as tax bases for the wage tax and the sales taxes, are biased in opposite directions. Whereas wage growth is underestimated by -0.38 percentage points, the growth of private consumption is overestimated by 0.42 percentage points.

Table 2.2: Test on Unbiasedness of Tax Base Forecasts

	Nominal GDP	Wages	Corporate Income	Private Consump.
α_1^B	0.41**	-0.38***	1.34	0.42***
	(0.20)	(0.07)	(0.83)	(0.14)
Obs.	112	68	55	34
R^2	0.00	0.00	0.00	0.00

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Neither the elasticity of the wage tax with respect to gross wages and salaries nor the elasticity of the sales taxes with respect to private consumption are biased (see Table 2.3). Instead, the elasticity for the overall tax sum with respect to nominal GDP is underestimated by -0.12. Furthermore, the elasticities of two profit-related taxes are biased in the sample period. The elasticity of the business tax is overestimated by 1.42 and the elasticity of the corporate tax by 2.37. Furthermore, the elasticity of the energy tax is slightly underestimated. Our results for the elasticities cannot directly be compared to existing studies as the literature has not focused on this issue before. Breuer (2015), however, investigated the tax ratio (relation of tax sum to nominal GDP) and could not detect a bias after reunification; for the overall sample beginning in 1968 he finds an overestimation.

Table 2.3: Test on Unbiasedness of Elasticity Forecasts

	Tax Sum	Wage Tax	Income Tax	Business Tax	Corpor. Tax	Energy Tax	Sales Taxes
α_1^{ε}	-0.12***	-0.57	1.20	1.42***	2.37***	-0.12*	-0.02
	(0.04)	(0.77)	(1.03)	(0.42)	(0.85)	(0.07)	(0.03)
Obs.	112	66	55	55	55	112	34
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.2 Weak Rationality

We continue with a test on weak rationality by estimating the following panel model,

$$\mathrm{FE}^i_{h,t} = \alpha^i_2 + \beta^i_1 \widehat{y}^i_{h,t} + u^i_{2,h,t} \; , \tag{2.7}$$

with i again either representing tax revenues of each tax type ($i = r$), the tax bases ($i = B$), or the elasticities ($i = \varepsilon$); $\widehat{y}^i$ represents the forecast for the corresponding variable. As stated in Section 4, a weakly rational forecast implies $\alpha^i_2 = \beta^i_1 = 0$. A constant or a beta that significantly differs from zero can also be seen as a more robust indicator for bias detection than in the previous section on unbiasedness.

The results for the tax revenue forecasts are shown in Table 2.4. The forecasts for the overall tax sum, the wage tax, and the corporate tax are weakly rational according to our definition. In the other cases, either the constant (income tax, business tax and sales taxes) or both coefficients (energy tax) are non-zero. That means, the growth of the income (business) tax is underestimated by 3.41 (1.82) percentage points and the growth of the sales taxes are overestimated by 0.41 percentage points. Contrary, the energy tax is underestimated by 0.12 percentage points. Consequently, those forecasts do not fulfill weak rationality as defined in Section 2.4. The values for the R^2 are also very low. Compared to existing studies, our results for the overall tax sum are in line with Breuer (2015).

Table 2.4: Test on Weak Rationality of Tax Revenue Forecasts

	Tax Sum	Wage Tax	Income Tax	Business Tax	Corpor. Tax	Energy Tax	Sales Taxes
α_2^r	-0.19	0.48	-3.41*	-1.82***	1.85	-0.22*	0.41**
	(0.30)	(0.29)	(1.09)	(0.27)	(1.07)	(0.09)	(0.09)
β_1^r	0.07	0.05	0.06	0.12	-0.17	0.15**	0.01
	(0.09)	(0.09)	(0.15)	(0.12)	(0.12)	(0.04)	(0.02)
Obs.	112	112	112	112	105	112	112
R^2	0.01	0.01	0.00	0.02	0.02	0.08	0.00

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results for the macroeconomic projections are mixed (see Table 2.5). While the forecasts for nominal GDP are weakly rational, the forecasts for gross salaries and wages, corporate and investment income and for private consumption are not. Growth of gross salaries and wages as well as of private consumption are underestimated by 1.08 resp. 1.78 percentage points and the corporate and investment income is overestimated by 1.15 percentage points. While the explanatory power of the forecast for corporate investment income and a constant remains very low for the variation of its forecast error, the opposite is true for private consumption. There, 45% of the variance of the forecast error can be explained by our model for weak rationality. Its large explanatory power supports our finding that the forecasts for private consumption are not weakly rational. Breuer (2015) also finds that nominal GDP forecasts are weakly rational.

Table 2.5: Test on Weak Rationality of Tax Base Forecasts

	Nominal GDP	Wages	Corporate Income	Private Consump.
α_2^B	-0.41	-1.08*	1.15**	-1.78*
	(0.44)	(0.45)	(0.32)	(0.58)
β_1^B	0.25	0.27	0.06	0.70**
	(0.13)	(0.18)	(0.10)	(0.19)
Obs.	112	68	55	34
R^2	0.08	0.10	0.01	0.45

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regarding the forecasted tax elasticities, the wage tax elasticity as well as the elasticities for the profit related taxes, which are income tax, business tax and corporate tax, are not weakly rational (see Table 2.6). Furthermore, the elasticity forecasts of the profit-related taxes and a constant explain between 46%, 52% and

56% of the variation of their forecast error. Again, this large explanatory power of a constant and the forecast support our finding of a biased forecast for each of the profit-related taxes. The wrong forecast for the elasticity by itself could be responsible for the complete tax forecast errors.

Table 2.6: Test on Weak Rationality of Elasticity Forecasts

	Tax Sum	Wage Tax	Income Tax	Business Tax	Corpor. Tax	Energy Tax	Sales Taxes
α_2^ε	-0.35*	-3.21*	-1.75***	-0.21	-1.79**	-0.08	-0.60
	(0.12)	(1.18)	(0.14)	(0.19)	(0.40)	(0.05)	(0.27)
β_1^ε	0.25	2.36	0.99***	1.14***	1.15***	-0.07	0.49
	(0.13)	(1.06)	(0.05)	(0.14)	(0.11)	(0.09)	(0.23)
Obs.	112	66	55	55	55	112	34
R^2	0.06	0.19	0.56	0.46	0.52	0.00	0.21

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.3 Driving Forces of Forecast Errors

Finally, we ask whether the forecast errors for the tax bases, the elasticities or both explain the errors of the tax revenue forecasts and how much both components contribute to the explained variation in these errors.[11] For this reason, we estimate the following panel model already introduced in Equation (2.5):

$$\mathrm{FE}_{h,t}^{r} = \beta_2^r \mathrm{FE}_{h,t}^{B} + \beta_3^r \mathrm{FE}_{h,t}^{\varepsilon} + \alpha_3^r + u_{3,h,t}^r \,.$$

If the tax forecast error can neither be traced back to the forecast error of the tax base nor the elasticity, this reads as $\beta_2^r = \beta_3^r = 0$ in our model. Besides, this implies a low value for the R^2. A large value for R^2 means that the forecast error of the tax can be explained to a high degree by the forecast error of the tax base, the forecast error of the elasticity and a constant. Ideally, these variables should have no explanatory power at all as would be indicated by values of the variables, that are statistically not different from zero. A significant constant could be interpreted as biasedness that cannot be assigned to the tax base and the elasticity while controlling for effects of the forecast error of the tax base and the forecast error of the elasticity.

The upper panel of Table 2.7 shows the results of the error-explanation-estimation. The overall tax forecast error is strongly determined by the forecast errors of both its tax base and its elasticity. A forecast error of one percentage point for the growth forecast for nominal GDP results in a forecast error of 1.15 percentage points for the growth forecast for the overall tax sum. If the elasticity is over-/ underestimated by one, this translates, c.p., to a forecast error of 1.45 percentage points. In combination with the constant, 81% of the variation in tax forecast errors is explained. Also, most of the analyzed individual tax forecast errors can

[11]To prevent running into issues of multicollinearity, Appendix B.1 contains the correlation coefficients between both forecast errors. There is no indication that our estimation results are blurred by a high correlation across the regressors.

be traced back to its tax base and/ or its linking elasticity. So, the forecast error of the wage tax is determined by the forecast error of the wages and of the linking elasticity, that is, the elasticity of the wage tax with respect to the sum of gross wages and salaries. A forecast error in wage growth of one percentage points leads to a forecast error of 1.90 percentage points of the wage tax. An elasticity forecast error of one increases the forecast error by 0.04 percentage points. In addition with the constant, this covers 67% of the wage tax forecast error's variation.

Table 2.7: Origin of the Tax Forecast Error and Relative Explanatory Power

	Tax Sum	Wage Tax	Income Tax	Business Tax	Corpor. Tax	Energy Tax	Sales Taxes
β_2^r	1.15***	1.90***	1.38**	0.69*	2.65**	0.28*	0.75***
	(0.09)	(0.10)	(0.38)	(0.27)	(0.62)	(0.09)	(0.10)
β_3^r	1.45***	0.04**	-0.24	-0.01	0.18*	1.46***	2.47***
	(0.07)	(0.01)	(0.14)	(0.12)	(0.06)	(0.11)	(0.18)
α_3^r	-0.26***	0.81***	-9.07***	-2.72***	-6.73***	0.17**	0.08
	(0.04)	(0.04)	(0.36)	(0.19)	(0.87)	(0.05)	(0.05)
Obs.	112	66	55	55	55	112	34
R^2	0.81	0.67	0.24	0.31	0.42	0.44	0.84
				Relative Explanatory Power			
FE^B	69%	93%	99%	100%	97%	6%	35%
FE^ε	31%	7%	1%	0%	3%	94%	65%

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The income tax forecast error can be explained by the forecast error of the profits but not significantly by the elasticity. An over-/ underestimation of the growth rate of corporate and investment income results in a forecast error of 1.38 percentage points of the growth of the assessed income tax. The relative low value for the R^2 of 24% reveals that there must be much more to the picture of the income tax forecast errors than simply a wrong prediction of the tax base. As for all profit-related taxes (income tax, business tax and corporate tax), the proper estimation of the assessment timeline might be another source of forecast error irrespective of the size and dynamics of the profits themselves. Gebhardt (2001) points to this source of forecast error. We, however, cannot investigate this channel as we have no access to the assessment timelines of the past. The forecast error of the business tax can weakly be explained by the forecast errors of its tax base but not by its elasticity. Consequently, the explanation of the variance of the forecast error of our model is just 31%. The forecast error of the corporate tax is significantly influenced by the forecast error of corporate and investment income and only weakly by the elasticity. If the corporate income growth exhibits a forecast error of one percentage point, this results in a forecast error of 2.65

percentage points of the corporate tax; a wrong prediction of the elasticity results in a forecast error of 0.18 percentage points. In combination with the constant, 42% of the variance of the forecast error of the corporate tax can be explained.

The forecast error of the energy tax is strongly determined by the forecast error of its elasticity and by a much weaker extent by its tax base. An over-/ underestimation of the elasticity of magnitude one increases the forecast error of the growth rate of the energy tax by 1.46 percentage points. Adding the constant, this leads to an R^2 of 44%. For sales taxes, 84% of the variation of the forecast error can be explained by the forecast errors of private consumption and the linking elasticity, which are both highly significant. So, a forecast error for the growth rate of private consumption of one percentage point leads to a forecast error of the growth rate of the sales taxes by 0.75 percentage points. A forecast error of the elasticity of one results in a forecast error of 2.47 percentage points of the growth rate of the sales taxes.

The analysis of the estimation results indicates only whether the forecast errors of the tax base or the elasticity are significant drivers of the tax forecast error. An immediately adjacent question is: How much does each of those two factors explain the tax forecast error? Or different: Are the tax forecast errors more driven by wrong predictions of the tax base or the elasticity? A mere look at the coefficients only reveals the influence of a standardized amount of the forecast error of the tax or the elasticity. It does not yet reveal how much it contributes to the overall magnitude of the tax forecast error. A helpful instrument to investigate how much both sources contribute to the tax forecast error is the squared semi-partial correlation coefficient.[12] It shows how much the R^2 changes when a coefficient is not included into the regression. Hence, we can interpret the squared semi-partial correlation coefficient of a variable as the additional explanatory power that is supplied by this additional variable. We put the squared semi-partial correlation coefficients of the tax base and the elasticity in relation to each other such that they sum up to one. Hence, the values in the lower panel of Table 2.7 illustrate which share of the R^2 that is explained by the tax base forecast error and the elasticity forecast error can be attributed to each of them. We call this the relative explanatory power; Appendix B.2 contains the absolute values.

For the overall tax sum, 69% of the explanatory power can be assigned to forecast errors of nominal GDP. This means the forecast error of nominal GDP had approximately twice as much influence on the tax forecast error variation than the elasticity in our sample period. Regarding the wage tax, we get an even clearer picture: 93% are attributed to the forecast error of the wages and only a small share to the elasticity. When looking at the profit-related taxes, nearly all the

[12]The semi-partial correlation of x and y, when holding z constant for y is defined as $r_{x(y.z)} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{1 - r_{yz}^2}}$, where the correlation between two variables i and j is defined as r_{ij}.

share is assigned to the forecast error of the corporate and investment income. However, the R^2 of the estimations are rather low compared to the other taxes.[13] As mentioned before, these are assessed taxes. The assessment timeline might complicate the tax forecasts (Gebhardt, 2001). A bad estimate for the assessment timeline can drive the forecast error of the tax forecast while not influencing the forecast error of the tax base or of the elasticity. For the energy tax, the picture is turned upside down. 94% of the share can be assigned to the forecast error of the elasticity. Again, this result is not surprising when looking at the estimation results. For the sales taxes we observe the largest value for the R^2. Both, the tax base and the elasticity, were significant although to a different degree. When looking at the squared semi-partial correlation coefficients, we get a clearer picture regarding their importance. 65% of the variance of the tax forecast error can be explained by the elasticity. Only 35% can be attributed to the forecast error of private consumption.

Overall, we find heterogeneous influence across the tax types which source of influence matters most for the tax forecast errors. Thus, less biased future tax revenue forecasts might be either achieved by more precise macroeconomic projections that are an exogenous input for the Working Party, methodological improvements or both. Whereas we stick to the methodological issue in the conclusion to this paper, we compare the accuracy of the government's macroeconomic forecasts to the performance of an independent institution in the following.

2.6 The Independent Forecaster

According to Lehmann and Wollmershäuser (2020) and Büttner and Kauder (2010) independent institutions publish better forecasts than the government. Even more specific, the macroeconomic projections of the German government for nominal GDP are outperformed by the forecasts of the Joint Economic Forecast (JEF) according to Lehmann and Wollmershäuser (2020). The JEF is a biannual consensus forecast of Germany's main Economic Research Institutes, which serves as the central input for the government's own macroeconomic projections. Moreover, the legal position of the JEF tremendously changed in 2017 as the German Federal Government finally implemented the Two-Pack Regulation No. 473/2013 of the European Parliament and the Council of the European Union. This regulation prescribes that national budgetary processes shall be enhanced or produced by an independent body or institution. An ordinance that came into effect on July 1, 2018 appointed the JEF to be the independent body that assesses and confirms the government's macroeconomic projections. Given the existent literature, one can

[13] This can also be seen in their low values for the explanatory power in Table B.1.

suspect the JEF might also outperform the WPTRE in its tax forecasts. Contrary to the WPTRE forecasts, which is based on accrual accounting just as the government's budget, the tax forecasts of the JEF are in terms of the System of National Accounts (SNA). That means, they cannot be compared to each other. Therefore, we compare the forecasts for the tax bases since a better tax base forecast should, c.p., result in a better tax revenue forecast. Büttner and Kauder (2010) argue in quite the same direction that better tax base forecasts should lead to better revenue predictions of various tax types and thus the overall tax sum. Breuer (2015) also suspects the WPTRE would produce better forecasts by using the macroeconomic projections of the JEF. We investigate these claims in the following and begin our investigation with a simple comparison of the forecast errors as well as the forecasts themselves and proceed with an analysis on statistical significance of the forecast deviations between the two institutions.

Table 2.8 presents the mean absolute forecast errors, $\text{MAFE}^{j,B} = 1/T \sum_T (|\text{FE}^{j,B}|)$ with $j = \text{JEF}, \text{GOV}$, and the mean absolute forecast deviation of the forecasts, $\text{MAFD}^B = 1/T \sum_T (|\widehat{y}^{B,\text{JEF}} - \widehat{y}^{B,\text{GOV}}|)$. We can confirm the finding of Lehmann and Wollmershäuser (2020) that the MAFE for nominal GDP by the JEF (1.14 percentage points) is smaller than the one by the government (1.19 percentage points). We observe a similar picture for the wage and private consumption forecasts, where the JEF outperforms the government by 0.07 and 0.10 percentage points, respectively. Only for the corporate and investment income, the forecasts by the government perform better than the ones by the JEF; the MAFE of the government is 0.18 percentage points smaller.

Table 2.8: Forecast Performance and Deviation

	Nominal GDP	Wages	Corporate Income	Private Consump.
MAFE^{GOV}	1.19	0.97	4.14	0.58
MAFE^{JEF}	1.14	0.90	4.32	0.48
MAFD	0.28	0.30	1.26	0.29
Obs.	105	68	55	34

Notes: The errors and the deviation are calculated for all four forecast horizons and are displayed in percentage points. The period under investigation runs from 1992-2019 for GDP. For the remaining tax bases it runs from 2001-2019 for the first two horizons and 2002-2019 for the last two horizons.

Instead of investigating the forecast errors, the mean absolute forecast deviation directly compares the growth forecasts of both institutions with each other. For nominal GDP, wages and private consumption, this deviation is around 0.30 percentage points. For corporate and investment income, the most volatile tax base, it is 1.26 percentage points. However, the absolute values give no information on whether and in which direction the government systematically deviates, on average, from the forecasts of the JEF. To test this, we run the following panel regression:

$$\widehat{y}_{h,t}^{B,\text{GOV}} - \widehat{y}_{h,t}^{B,\text{JEF}} = \text{DEV}_{h,t}^{B} = \alpha_4^B + u_{4,h,t}^B \,.$$

The results to this estimation are displayed in Table 2.9. The forecasts for private consumption by the government are 0.21 percentage points larger than the ones by the JEF. For the remaining tax bases, the government's forecasts do not seem to significantly deviate from the JEF's predictions; however, the average deviations are, with the exception of nominal GDP, not small at all. In Table 2.2 in Section 2.5, we observed an overestimation of 0.42 percentage points for private consumption in the government forecasts. Table 2.9 shows that the forecasts for private consumption by the JEF are on average 0.21 percentage points smaller. Consequently, using the JEF instead of the government forecast for private consumption would, c.p., remove or at least reduce the bias for this variable. Therefore, the bias for the sales taxes (see Table 2.1) should vanish or at least shrink.

Table 2.9: Test on Systematic Forecast Deviations

	Nominal GDP	Wages	Corporate Income	Private Consump.
α_4^B	-0.03	-0.11	0.18	0.21***
	(0.03)	(0.08)	(0.45)	(0.08)
Obs.	105	68	55	34
R^2	0.00	0.00	0.00	0.00

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Despite the fact that the government does not seem to systematically deviate—with the exception of private consumption—from the JEF, the deviations are quite large as shown in Table 2.8. These deviations might be triggered by an informational advantage of the government compared to the JEF. Usually, the government publishes its macroeconomic forecasts one or two weeks after the JEF publications. If the government incorporates new information effectively and thus deviates from the JEF, the precision of the government's forecasts should increase. Frankel and Schreger (2016) suggest to run the following regression:

$$\text{FE}_{h,t}^{B,\text{GOV}} = \alpha_5^B + \beta_4^B \text{DEV}_{h,t}^B + u_{5,h,t}^B \,. \tag{2.8}$$

Table 2.10 shows the corresponding estimation results. We indeed find that the government's forecast errors increase by the deviation from the independent forecaster as the positive β_4^B coefficients suggest. However, the effect is only statistical significant for nominal GDP and private consumption. The informational advantage the government possibly faces does not result in a higher accuracy of their macroeconomic projections. Rather the opposite holds true as the accuracy worsens, underpinning the larger overall absolute deviations. Furthermore, all tax base forecasts show systematic overall biases readily observable by the estimated constant terms α_5^B.

Altogether, we can support the hypothesis that the independent forecaster JEF can improve the tax forecasts, especially through the channel of more accurate macroeconomic projections. We find a significant improvement for nominal GDP and private consumption as tax bases. For the other two tax bases, the JEF forecasts are at least as good or better than the forecast by the government despite the information advantage of one or two weeks when it publishes its forecast. In the end we can ask why the government produces macroeconomic forecasts at all. The forecasts can be done at least as good by the independent institution. Moreover, according to the fiscal rules in the EU, the government forecast needs to be evaluated or even produced by an independent institution. Regarding our

Table 2.10: Test on Informational Advantage of the Government

	Nominal GDP	Wages	Corporate Income	Private Consump.
α_5^B	0.37***	-0.32***	1.27*	0.77***
	(0.14)	(0.07)	(0.77)	(0.27)
β_4^B	0.77*	0.50	0.36	0.27*
	(0.46)	(0.45)	(0.41)	(0.16)
Obs.	105	68	55	34
R^2	0.04	0.04	0.02	0.34

Notes: Clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

findings, replacing the government forecast with the JEF forecast would increase the accuracy and, on top, save resources that are used to produce the government's tax base forecasts.

2.7 Conclusion

This paper disentangles tax revenue forecast errors into two major sources: a wrong assessment of the macroeconomic outlook and a false prediction of the elasticity linking the tax base to its corresponding tax type. The analysis for six major tax types and the overall tax sum of Germany reveals that both sources matter for biases in tax revenue forecasts. However, we observe a heterogeneous degree of relative importance of both sources across tax types. Whereas more than 90% of the explained forecast error of the profit-related taxes (income tax, business tax, corporate tax) as well as the wage tax can be attributed to wrong macroeconomic assumptions, the opposite holds true for the two remaining taxes. Here, 94% (energy tax) or two-thirds (sales taxes) of the explained revenue forecast errors can be traced back to wrong assumptions on the linking elasticity. For the overall tax sum, 69% of the error can be attributed to wrong macroeconomic predictions and 31% to false assumptions on the elasticity. Especially the errors due to wrong macroeconomic projections can be mitigated according to our results. The comparison of the government's macroeconomic projections with those of an independent forecasting institution reveal that the independent forecaster produces better forecasts for nominal GDP and private consumption and at least as good predictions for gross wages and corporate income.

Our results give rise to improvements of the underlying tax revenue forecasting process that in the end might go hand in hand with even more precise forecasts. First, the macroeconomic projections should be fully outsourced to the independent institution. According to our results, this would reduce the bias introduced by the first source and might also save resources within the ministries that are used to produce the macroeconomic forecasts. Second, it is also likely that methodological improvements reduce future tax revenue forecast errors. Whereas the academic literature on macroeconomic forecasting, and here especially GDP, is flourishing, only a handful of newer articles exist that transfer modern forecasting techniques to issues of fiscal forecasting (see, for example, Asimakopoulos *et al.*, 2020). Dynamic factor models or mixed-frequency vector autoregressions are the workhorses in applied macroeconomic forecasting and have proven to outperform other techniques. Furthermore, new and larger data sets can be explored to further improve existing techniques such as elasticity methods. In the end, there is room for improvement to reduce forecast errors so that the necessity of tax revenues to serve as guidelines for future expenditures is strengthened, preventing governments to run into excessive public debts.

Chapter 3

The State-Dependent Effects of Social Security Contributions on the Macroeconomy

3.1 Introduction

The Corona crisis led to a need of fiscal policy interventions even greater than already known by the Great Depression. By the end of 2021, the USA have spent already more than four trn. Dollar in response to COVID-19.[1] The EU set up the largest stimulus package in its history, the NextGenerationEU package, with a volume of more than two trn. Euro.[2] So, it is more important than ever to understand how government spending and revenues influence the economy. Moreover, we need to know whether certain conditions like the position in the business cycle affect the outcome.

There is already a vast amount of literature on fiscal multipliers building on the seminal work of Blanchard and Perotti (2002).[3] Especially the literature on spending multipliers is quite large. We have literature on how different branches of spending, various estimation methods and even how different data adjustments like the chosen SNA standard influence the size of the multiplier (e.g. Gechert, 2015; Capek and Crespo Cuaresma, 2020). On the revenue side, the evidence is more limited and concentrated on tax multipliers (e.g. Chahrour *et al.*, 2012; Favero and Giavazzi, 2012; Christofzik *et al.*, 2022). However, the revenue side of the government sector contains more than only taxes. There is a second large pillar, it is built

[1]https://www.usaspending.gov/disaster/covid-19.

[2]https://ec.europa.eu/info/strategy/recovery-plan-europe_en.

[3]Blanchard and Perotti (2002) represent the modern branch of quantifying fiscal multipliers. Former ones can be found in e.g. Evans (1969).

on: social security contributions. Capek *et al.* (2021) have shown that adjusting for social security contributions alters the tax multiplier quantitatively. There is some evidence that the economic agents show heterogeneous responses to taxes on labor and social contributions in the microeconomic literature (e.g. Saez *et al.*, 2012). According to my knowledge, Gechert *et al.* (2021) are the only ones estimating social security contributions multipliers separately. Their work concentrates on Germany and provides also estimations for social benefits. However, they do not check for possible differences in the multiplier when considering the position in the business cycle. In the literature on spending multipliers, there is a large branch that deals with this question and comes to different results. Whereas e.g. Auerbach and Gorodnichenko (2012) or Caggiano *et al.* (2015) came to the conclusion, that the multipliers are larger in recessions, others like Ramey and Zubairy (2018) or Owyang *et al.* (2013) found no difference. Regarding tax multipliers, Sims and Wolff (2018) have shown various characteristics of state-dependence that also include the position in the business cycle. But for social contributions, it is unknown if they influence the macroeconomy differently when comparing boom and bust.

This paper tries to shed some light to this question by using Local Projections in a panel setting similar to Alpanda *et al.* (2021). This method has already often been used, especially for investigating possible state-dependence of fiscal multipliers (e.g. Auerbach and Gorodnichenko, 2013; Ramey and Zubairy, 2018). To identify the shocks, I use a narrative approach as introduced by Romer and Romer (2010). More precisely, I rely on forecast errors similar to Auerbach and Gorodnichenko (2012) or Auerbach and Gorodnichenko (2013). My panel consists of Canada, Japan, the United Kingdom (UK) and the United States of America (USA). Due to data limitations, the analysis is restricted to those four countries. I estimate the effects of an unexpected shock of the growth rate of social security contributions to GDP, the CPI, the GDP deflator, private consumption, capital formation, the wages and the unemployment rate. In a first step, this estimation is applied linearly. In a second one, the model is adapted to a state-dependent version that takes into account the position in the business cycle. I distinguish between the recession as one state and non-recessionary periods as the other state.

In the present paper, I find a positive influence of a reduction in social security contributions on real GDP in beginning followed by a negative period. The effects are very small and primarily driven by the stimulation of private consumption and capital formation. Considering the microeconomic literature, it is not surprising that the effects are smaller when compared to tax changes. A social security contribution has an insurance component, which the employees seem to be aware of Saez *et al.* (2012). The price level is increased and there is an instant reduction of the unemployment rate, that decreases in the subsequent quarters. Most of the series reveal an ambiguous shape that can be explained when controlling for the position in the business cycle, where I observe a pro-cyclical response: The effect on GDP is positive during a non-recessionary period and negative in a recession. This helps to explain the negative periods in the linear model. Again, the effect is driven by private consumption and capital formation. Although the effect on prices is mostly positive, it turns negative especially during a recession for some time. I only observe a lowering of the wages by the reduction of the tax wedge during a recession. But in the remaining periods, we have an additionally positive effect on wages that slightly increases over time. The effect on the unemployment rate is weak in both states. However, there seems to be the only one counter-cyclical response after twelve quarters with a small decrease in a recession and a small increase in the remaining periods. A possible explanation for the negative reactions of private consumption and capital formation could be a forward-looking behavior of the economic agents. They might be aware of their position in a recession and might assume the government would react with a revenue increase in the future to the current reduction in social security contributions. Most of my data set is in a period that is influenced by the Zero Lower Bound. Sims and Wolff (2018) found lower tax multipliers for those time periods. The effects on social security contributions multiplier could be similar. Consequently, I it is possible that I underestimate the effects of social security contributions on other macroeconomic variables in normal times.

These results are mostly in line with Gechert *et al.* (2021). However, Gechert *et al.* (2021) explore the effects of a change in social contributions in Germany, whereas I observe the results in a panel consisting of Canada, Japan, the UK and the USA. The pro-cyclical response to the shock in social security contributions corresponds to the pro-cyclical tax multipliers of Sims and Wolff (2018).

The remainder of this paper is organized as follows. Section 3.2 describes the data set and the methodological approach chosen in this paper. Section 3.3 contains the findings starting with the linear model and continuing with the state-dependent one. Section 3.4 concludes.

3.2 Empirical Strategy

3.2.1 Data Set

Similar to Auerbach and Gorodnichenko (2013), I use data provided by the OECD on their member countries. The OECD regularly publishes its Economic Outlook (EO). Together with the IMF forecast, the World Economic Outlook (WEO), it can be seen as the most important economic forecast for policymakers worldwide. Both forecasts have a high quality and credibility. However, the EO has two major advantages compared to the WEO: First, the EO contains quarterly data whereas the WEO publishes yearly data. Second, the EO has a vast amount of forecasted variables. It does not only include the basic variables on the macroeconomy like GDP, the unemployment rate or the output gap. It also includes detailed subcategories for the sector accounting. This level of detail can usually only be found in local forecasts of countries in their own language performed by their political or research institutions as e.g. the Greenbook forecasts by the Federal Reserve in the USA or the Joint Economic Forecast by economic research institutes in Germany. Since the OECD is one of the global leading forecasters, we can assume, they include every planned policy change into their forecast. According to Auerbach and Gorodnichenko (2013), the OECD uses its local presence in the member countries to gain specific knowledge for the forecasts for each country. Furthermore, Auerbach and Gorodnichenko (2013) explain, the OECD discusses their forecasts in great detail with local policymakers and government experts.[4] The quality and performance of the forecast is as good as benchmarks from the private sector as Lenain (2002) and Vogel (2007) show.[5] Furthermore, the data provided by the OECD has the advantage of relying on a unified methodology, which is important to easily include it into the a estimation.

In this large amount of variables, the social security contributions received by the government are also included as a separate variable. That particular variable covers every social security payment made to the government regardless if it was paid by the employer (the firm sector) or the employee (the household sector). Hence, although I am not able to distinguish the effects of a shift in the bearing of the social security contribution, I can observe the effects of an increase in the social security contributions per se - no matter if born by employees or firms.

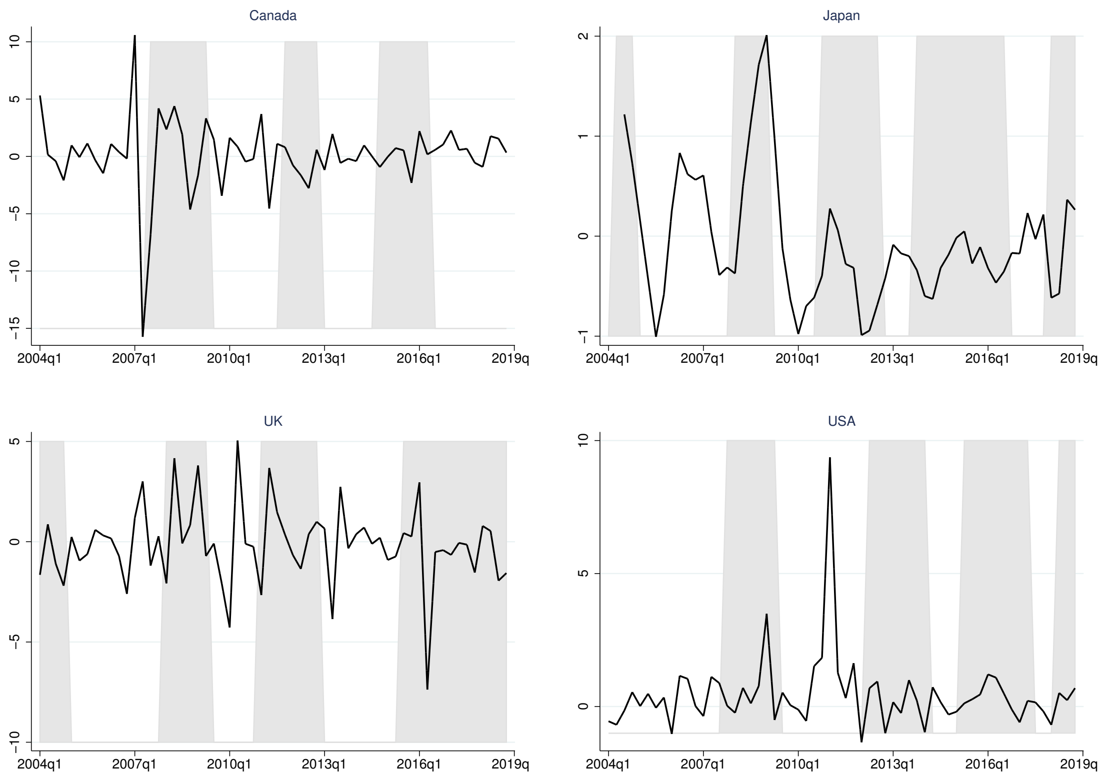
In this paper, I follow Romer and Romer (2010) by using a narrative approach. I use the EO forecasts to obtain forecast errors similarly to Auerbach and Gorodnichenko (2012) as well as Auerbach and Gorodnichenko (2013) to investigate the effects of a change in the amount of social security contributions. The forecast error is based on the quarter over quarter growth rate of the social security contributions received by the general government. The realized growth rate for the social security contributions is subtracted from the most recent forecast for quarter q that is prior to this quarter q to construct the forecast error. Consequently, a positive value for the forecast error can be interpreted as a surprise social security contributions shock that reduces the contributions. Vice versa, a negative value corresponds to a shock that raises the contributions. According to my knowledge, I am the first to exploit the EO forecast errors of the social security contributions

[4]Auerbach and Gorodnichenko (2013) call this acquiring of local knowledge "reality checks".

[5]Additional information regarding the EO can be found online at http://www.oecd.org/faq/0,3433,en_2649_33733_1798284_1_1_1_1,00.html.

to estimate its effects on other macroeconomic variables. Figure 3.1 displays the shock series.

Figure 3.1: Exogenous shocks to social security contributions (in percentage points)



Notes: Exogenous shocks to social security contributions in percentage points. The figure shows the forecast error of the OECD EO for the growth rate of social security contributions received by the general government (in percentage points) for the period between 2004q1 and 2018q4. A positive signs indicates an expansionary shock (cut in social security contributions). A value of one means that the growth rate of nominal security contributions has been overestimated by one percentage point. The shaded areas indicate recessionary periods. A recessionary month is a period, where the OECD recession indicator identifies the economy to be in a recession. Whenever the OECD rates the majority of the months within a quarter as a recession, I treat the quarter as a recession.

The data spans from the first quarter of 2004 to the final quarter of 2018. Unfortunately, there exists no usable quarterly data before 2004. On the other hand, there are restrictions at the end of the time series due to the corona crisis. One could suspect small distortions in the last quarter of 2019 with the beginning spread of the virus. Moreover, the forecasts in the year 2020 were highly driven by the corona crisis itself. They were restricted to a smaller variety of variables and many forecasters used different methods than in normal times. The panel consists of four countries: Canada, Japan, the UK and the USA.

In a second step, the analysis of this paper is refined by taking into account the position in the business cycle. Again, it is important to have a database that consistently defines the states across the countries. Additionally, the methodology should be as similar as possible to the one used by the OECD to minimize the statistical noise of the estimation results and prevent the estimation from possible misspecification. To match with these criteria, I chose the OECD recession indicator provided by the Federal Reserve. The data can be accessed via their statistical database Federal Reserve Economic Data (FRED).[6] On the basis of composite leading indicators, the OECD identifies turning points of the business cycle. Whenever the economy is identified to be in a month of a recession, the indicator takes the value 1 and 0 otherwise. The published indicator is refreshed on a daily basis. More details regarding the indicator can be found online. I aggregated the monthly indicator to a quarterly dummy series. Whenever the majority of the months indicated a recession, I identified the quarter as a recession. The remaining quarters are classified as non-recessionary periods.

Next to real GDP, I also analyze the effects on the GDP deflator, the CPI, real private consumption, nominal wages and the unemployment rate.

Graphs that display the time series of all variables for all countries can be found in Appendix C.1.

3.2.2 Empirical Model

Similar to Ramey and Zubairy (2018) as well as Auerbach and Gorodnichenko (2013), I use a Local Projection Method to estimate the Local Impulse Response Functions (LIRFs) as proposed in Jorda (2005). For each variable and horizon a separate regression is used that does not depend on the other ones. That makes the Jorda-Method more flexible and easier to handle as has been highlighted i.a. by Jorda (2005), Stock and Watson (2007), Auerbach and Gorodnichenko (2013), Ramey and Zubairy (2018) or Alpanda *et al.* (2021). It is less prone to misspecification since the shape of the impulse responses is not constrained as in a Structural Vector Autoregression (SVAR) model. Since not all variables have to be included in every equation, it allows slimmer equations than an SVAR model and allows to focus on a single variable more easily. Moreover, the left-hand-side variable does not have to enter the equation in the same form as the right-hand-side

[6] The links to the used time series:
https://fred.stlouisfed.org/series/GBRRECDM,
https://fred.stlouisfed.org/series/USARECDM,
https://fred.stlouisfed.org/series/CANRECDM,
https://fred.stlouisfed.org/series/JPNRECDM.

variables. This allows a straightforward analysis of the empirical relationship of macroeconomic variables of different forms as the unemployment rate (expressed in percentage points) and GDP (expressed via a log-transformation). Plagborg-Moller and Wolf (2020) demonstrate that a local projection and a VAR approach estimate the same impulse in linear frameworks.

On the other hand, the Jorda method has also its weaknesses as described in Ramey and Zubairy (2018). Since the horizons are estimated independently, LIRFs are usually more bumpy than the traditional IRFs. Furthermore, you need to have longer time series since with every additional horizon, an observation from the end of the sample is lost. A problem that can be especially severe in macroeconomics and critical when estimating long-run relationships of variables. Ramey (2012) compares the results of the Local Projection Method with the estimates of a standard VAR and a dynamic simulation (as in Romer and Romer, 2010) using military news shocks. The results for the Jorda-method are similar for the short-term horizons. For longer horizons, they are statistically different and show oscillations that do not occur in the estimations of the other two methods. Consequently, I restrict my analysis to a short horizon, which is also more important for business cycle analysis.

The equation for the linear panel model reads as follows:

$$x_{t+h,i} = \alpha_{h,i} + \psi_h(L)y_{t-1} + \beta_h \mathrm{FE}_{i,t} + \epsilon_{t+h,i}. \tag{3.1}$$

In x enter the different macroeconomic variables whose reactions we want to estimate. These are: real GDP, the GDP deflator, the CPI, real private consumption, nominal wages and the unemployment rate. y is a vector of control variables, where ψ_h is a polynomial in the lag operator. Like Auerbach and Gorodnichenko (2013), the left-hand side variable is included in the control variables as are the lagged social security contributions. Moreover, nominal wages and its lags are added as control variables.[7] $\psi(L)$ is a polynomial of order 4. FE is the forecast error of the social security contributions. All variables except of the unemployment rate and the forecast error enter the equation in logs. The impulse responses are constructed as a sequence of all the β_hs of the single regressions for each horizon. Similar to Auerbach and Gorodnichenko (2013) and Alpanda *et al.* (2021), I use a panel estimation, allowing the intercepts α to vary by country but constraining coefficients to be the same. The standard-errors (ϵ) are clustered on the country-level.[8]

State-dependent models can be easily implemented via Local Projections as has been highlighted i.a. in Auerbach and Gorodnichenko (2013), Ramey and Zubairy

[7]When wages enter the equation on the left hand side, the right hand side lag operator is reduced.

[8]This is similar to the implementation in Jorda (2005).

(2018) or Alpanda *et al.* (2021). This is particularly true for a panel framework. Moreover, there is no assumption needed on the duration of the state or on the mechanism of transition to the other regime. Similar to Ramey and Zubairy (2018) and Alpanda *et al.* (2021), I extend my model to a state-dependent version:

$$\begin{aligned} x_{t+h,i} = I_t[\alpha_{A,h,i} + \psi_{A,h}(L)y_{t-1} + \beta_{A,h}\mathrm{FE}_{i,t}] \\ +(1-I_t)[\alpha_{B,h,i} + \psi_{B,h}(L)y_{t-1} + \beta_{B,h}\mathrm{FE}_{i,t}] + \epsilon_{t+h,i}. \end{aligned} \tag{3.2}$$

Here, I is a dummy variable that hast the value one whenever the recession indicator hints to a recession and zero otherwise. Via this structure of the estimation equation, all of the variables can vary according to the state. The remaining structure is identical to the linear version.

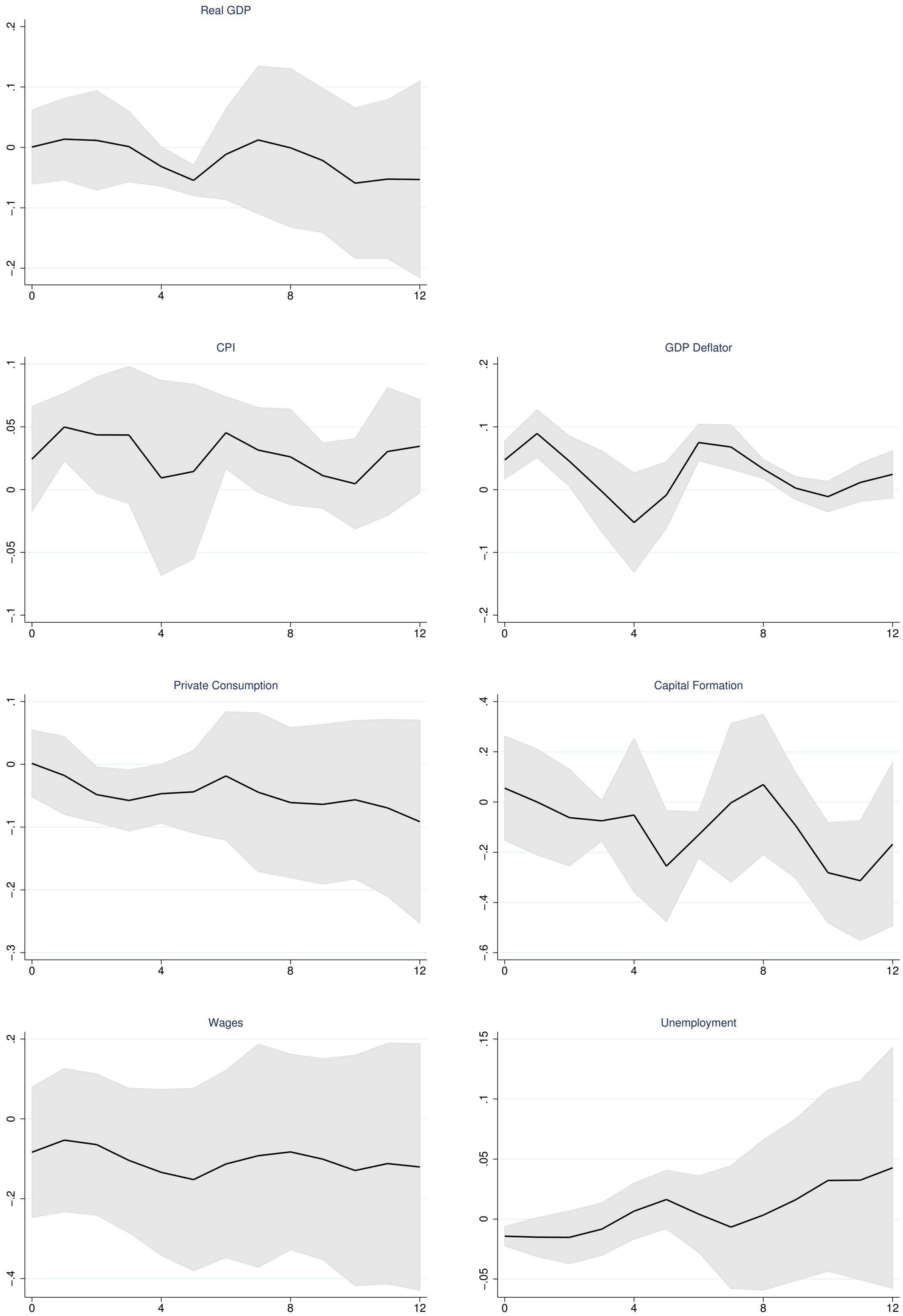
3.3 Results

3.3.1 Linear Model

The resulting values of the βs of equation (1) and equation (2) can be interpreted as semi-elasticities of the respective macroeconomic variable of the right hand-side with respect to the growth rate of the social security contributions received by the general government.

The baseline estimation is the linear model of equation (1). In the first quarters, GDP reacts slightly positive to a surprise reduction in social security contributions. E.g. it has a value of 0.02 after one quarter. That means a non-anticipated reduction of the growth rate of the social security contributions by one percentage point leads to a rise of the real GDP of 0.02 percentage points after one quarter. After a few quarters, it turns negative for a short period and switches back to positive values again (see Figure 3.2). However, in most of the quarters, the 90%-confidence interval reveals that these reactions are not different from zero. In Section 3.3.2, it can be seen that the bumpiness of the series is not exclusively caused by the estimation via Local Projections. The instability of the linear model can also be explained by the heterogeneous reactions depending on the macroeconomic environment. In a linear model, these states are mixed together in one estimation equation that cannot adequately deal with it. This is not only true for the effect on GDP but for most of the estimations. In a linear model, that is blind for the heterogeneity, this affects the estimations similar as in a model with an omitted variable bias. Gechert *et al.* (2021) also found a positive impact of a social security contributions cut on GDP, which turns negative after 1.5 years.

Figure 3.2: Linear response (in %)



Notes: Impulse response of specified variables to an exogenous shock equal to one percentage point of growth in social security contributions at 0-12 quarters after the shock. The shaded area shows the 90%-confidence interval.

The price reactions to a reduction of social security contributions are positive. I distinguish two variants of inflation: First, the inflation for the consumers, the CPI. Second, the inflation for the overall economic system, the GDP deflator. The effect on the CPI is positive for the entire observation period. But it turns permanently insignificant after six quarters with already a phase of insignificance after four quarters. The effect on the CPI hits its peak after one quarter with 0.05. So, an unexpected fall of the growth rate of the social security contributions by one percentage point induces a rise of the CPI by 0.05 percent after one quarter. After this immediate maximum, the effects slowly starts to phase out. However, there is a kink at the fourth and fifth quarter. After 12 quarters, the effect is only 0.02 and insignificant. For the GDP deflator, the effect is similar but larger in its size and less noisy as can be seen from the smaller confidence intervals. Again, the effect comes to its maximum very soon: After one quarter. Here, a sudden increase of the growth rate of the social security contributions by 1 percentage point leads to a 0.09 percent higher GDP deflator. After this, there is an even larger kink than observed for the CPI around the fourth quarter at which the effect also turns insignificant. After nine quarters, the effect is permanently insignificant and close to zero. Consequently, the estimated effect on the prices is only transitory. Gechert *et al.* (2021) concentrated for the price effect on the GDP deflator. They observed a positive effect of a cut in social security contributions on GDP, too. In their findings, we can also observe a dip in the IRF after a few quarters. However, it is less pronounced and a quarter earlier than in the present paper.

There is a small positive influence of lower social security contributions on private consumption that immediately declines and turns slightly negative in the second quarter. However, the effect turns slightly positive again around the sixth quarter. But for nearly the complete series the effects are statistically insignificant. Gechert *et al.* (2021) found a more pronounced positive effect of a cut in social security contributions on private consumption that turns negative after a longer time period (nine quarters). The ambiguous response will be explained in the next chapter by controlling for the position in the business cycle.

The effect on capital formation by a negative social security contributions shock is slightly positive in the beginning. However, the effect is mostly statistically insignificant on the 10%-level and the LIRF is quite bumpy in this non-state-dependent specification. The effect declines and turns negative after a few quarters with an exception around the eighth quarter. This result is somehow in contrast to Gechert *et al.* (2021) that found a positive effect for the first 14 quarters.

There is no statistically significant influence of the reduction in social security contributions on the nominal wage sum. However, there is a non-significant negative influence. Technically, the social security contributions are a part of the gross wage sum. So c.p., a reduction of the social security contributions means also a reduction of the gross wages. It is a mere reduction of the tax wedge as highlighted by Gechert *et al.* (2021). This effect should also be permanent. That fits exactly to the permanently negative influence we can see in Figure 3.2. Besides this technical part, there seems to be no effect of the social security contributions on the wages. Similarly to the findings in the present paper, Gechert *et al.* (2021) observe a reduction in the gross wage sum in the beginning. However, their series starts to rise after a few quarters and terminates in a positive effect on the wage sum.

There is a second dimension of the labor market that I cover: The effect on the unemployment rate. There is a very short significant impact of the social security contributions on the unemployment rate. An unexpected decline of the growth rate of the unemployment rate of one percentage point immediately reduces the unemployment rate by 0.02 percentage points. However, this effect is only weakly significant and vanishes very fast. In the subsequent quarters, the effect gets insignificant and declines. Hence, I cannot measure a permanent significant effect of a reduction of the social security contributions on the unemployment rate. Moreover, the significant effect in the beginning is rather negligible. Gechert *et al.* (2021) did not investigate the effect of social contributions on the unemployment rate but on employment. However, their finding of a positive influence of a negative social security contributions shock on employment is in line with my findings. The effects on the unemployment rate I find in the present paper return faster to zero than the employment rate in Gechert *et al.* (2021). Moreover, they did not observe a sign change as I do.

The effects found here, could be influenced by the Zero Lower Bound (ZLB) that is present in most of the time period I investigate. Christiano *et al.* (2011) found lower tax multipliers for these periods. It is possible that the ZLB has the same influence on the effects of changes in social security contributions.

To put the results of the linear model in a nutshell: There seem to be only small effects of an unexpected reduction in social security contributions on the economy except of a transitory higher inflation. However, that picture changes when the model accounts for the position in the business cycle.

3.1 reports the point estimators at three different horizons (four quarters, six quarters and eight quarters) for the components of the linear model. The numbers are provided to enhance comparability with other studies. However, the focus is on the graphical represantation of the LIRFs to give a representation that is easier accessible.

Table 3.1: Response to social security contributions shock at different horizons (in %)

	4 quarters	6 quarters	8 quarters
Real GDP	-0.03	-0.11	0.00
CPI	0.01	0.05*	0.03
GDP deflator	-0.05	-0.02	0.03**
Private consumption	-0.05	0.08	-0.06
Capital formation	-0.05	-0.13*	0.07
Wages	-0.13	-0.11	-0.08
Unemployment	0.01	0.00	0.00

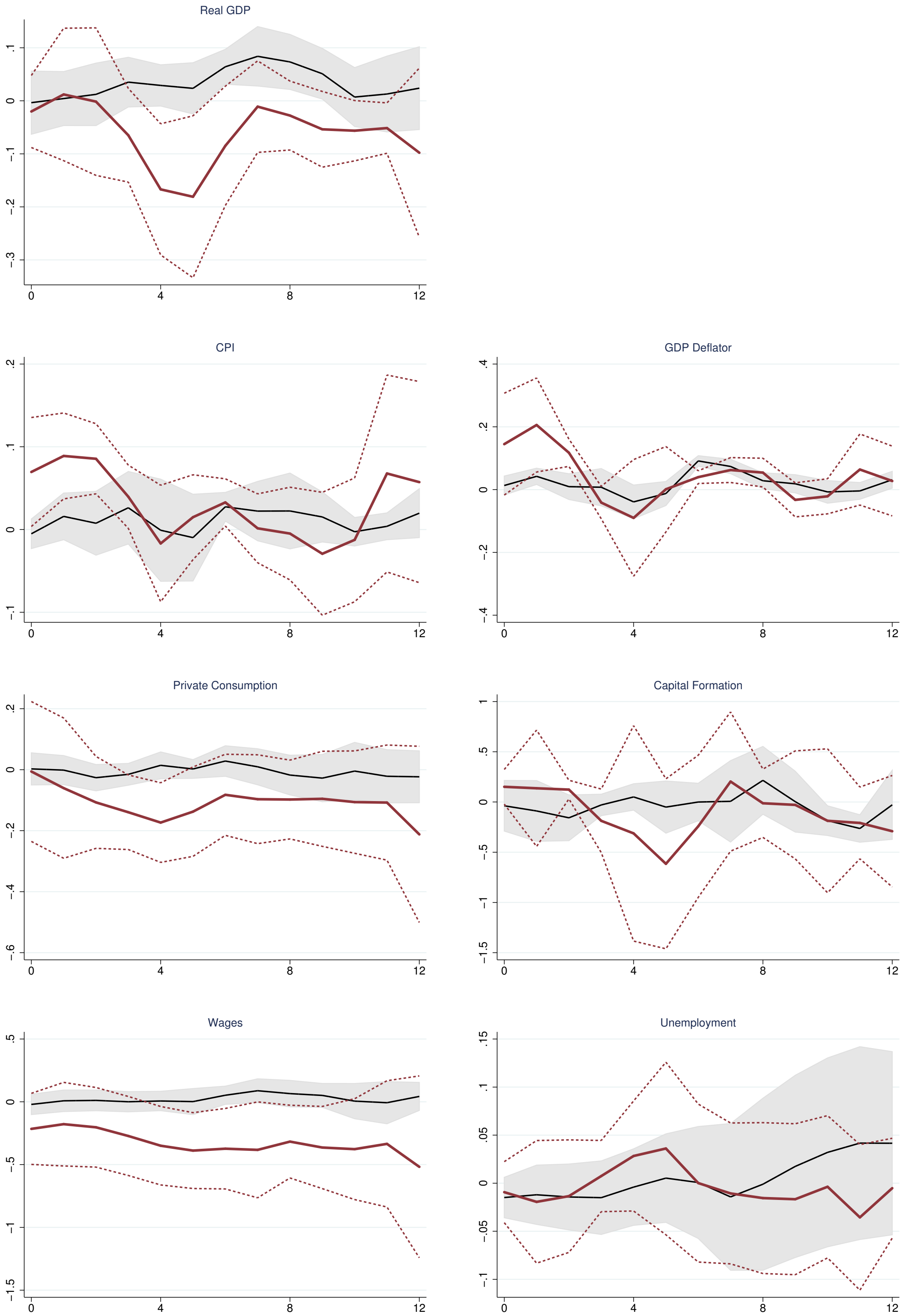
Notes: Level of significance: *** p<0.01, ** p<0.05, * p<0.1.

3.3.2 State-Dependent Model

When distinguishing between boom and bust, as is done in equation (2), I come to different results than in the previous subsection. More important, I can show there is some heterogeneity between both of the states.

Figure 3.3 displays the estimated LIRFs for the state-dependent cases. The thick maroon lines are the responses in the recessionary state. Their 90%-confidence intervals are represented by the maroon dashed lines around them. The estimates for the non-recessionary state are represented by the solid black line and their 90%-confidence intervals by the shaded area around it. In a recession, the reduction of social security contributions has only a very small and short-lived positive influence on real GDP in the beginning. While an unforeseen reduction in the growth of social security contributions by one percentage point leads to an increase of real GDP by 0.03 percent after one quarter, the effect turns negative after two quarters. Five quarters after the impact, the reaction reaches its lowest value with -0.18. As can be seen below, this is due to a reduction in private consumption. After seven quarters the effect gets close to zero. The reaction of real GDP does not change much in the following quarters. Most of the effects found for real GDP in a recession are not statistically significant at the 10%-level. However, the pattern of the reaction in a non-recessionary period is quite different. The effect is even smaller in the beginning with barely more than zero after one quarter. But it rises almost steadily until the seventh quarter, where the effect reaches its maximum value of 0.08. The positive effect for the sixth to the eighth quarter is also statistically significant at the 10%-level. After this peak, the effect declines to values which are close to zero and shows no sign of change any more after nine quarters. When looking at both disentangled series, one can see how the shape of the linear model is caused: The minimum of the recessionary series around the fifth quarter leads to a minimum in the linear model. Whereas, the moderate maximum of the expansionary series shapes a subsequent maximum in the linear model around the seventh quarter. These findings are in line with the ones of Sims and Wolff (2018) for the tax multiplier, where they also observed a pro-cyclical behavior. However, they do only observe a reduced size of the multiplier that does not induce a sign-change.

Figure 3.3: State-dependent response (in %)



Notes: Impulse response of specified variables to an exogenous shock equal to one percentage point of growth in social security contributions at 0-12 quarters after the shock. The maroon line indicates the response during a recession with the dashed lines indicating the 90%-confidence interval surrounding it. The black line indicates the response during a non-recessionary period. The shaded area displays its corresponding 90%-confidence interval.

There is also some heterogeneity in the reactions of the price level. As in the linear case, I look into the effects of the CPI and the GDP deflator. In a recession, my estimation shows an immediate effect of the consumer prices to a shock in social security contributions. The effect has its peak after one quarter with 0.09. This means that an unexpected drop in the growth of social security contributions by one percentage point leads to a rise in the CPI of 0.09 percent. In the following quarters, the effect declines and turns even negative with a local minimum at the fourth quarter. Then, there is a short-lived rise in the reaction for two quarters before a period of decline to a local minimum at the ninth quarter. Thereafter, the reactions phases out and gets close to zero. The remaining bumpiness of the series might be explained by the estimation method, the Local Projection Method. Except of the positive effect at the first and the second quarter after the shock, all effects are statistically insignificant. Again, the reaction of the price level estimated by the GDP deflator shows a similar behavior but larger in its size. During a recession, the GDP deflator immediately reacts with a positive response to a shock in social security contributions. The effect reaches its maximum with 0.20 after one quarter. Thereafter, the effect declines sharply and gets to a minimum of -0.09 after four quarters. In the subsequent quarters, the effect phases out and gets back to zero. Only the positive effects at the first and second quarter as well as the ones between the sixth and eigth quarter are statistically significant. In non-recessionary times, the price level reacts weaker to a social security contributions shock. The CPI has an overall increasing trend until the peak with 0.03 after six quarters. However, there is a dip between the third and sixth quarter. After the sixth quarter, the effect declines and vanishes. Only the reaction after the sixth quarter is statistically significant. The effect on the GDP deflator is similar but larger in its size. There is a short and weak maximum after the first quarter with 0.05. Thereafter the effect declines until the fourth quarter and turns even slightly negative. However it rises again to its maximum effect of 0.09 after six quarters. This peak at six quarters is equivalent in its timing to the effect on the CPI. After this maximum, the effect declines and gets close to zero. Contrary to the CPI, the effect on the GDP deflator is not only larger but also more significant. The positive effects after the first quarter and around the peak at the sixth quarter are statistically significant at the 10%-level resp. the 5%-level. On the one hand, the effects during expansions mitigate the effects of the linear model. On the other hand, the expansionary times shape the local maxima after six quarters in the linear model, whereas the strong reactions in the beginning and the subsequent drop in the linear model are caused by the recessionary times.

The heterogeneous patterns observed for real GDP can also be observed for real private consumption. In a recession, the effect starts slightly above zero and declines until the fourth quarter, where it has its minimum with -0.17. That means an unexpected drop in the growth rate of social security contributions of one percentage point reduces private consumption by 0.17 percent. Subsequently, the effect gets closer and closer to zero with an exception after twelve quarters, where a drop can be observed. However, few of these reactions are significant. Although, there are weaker but clearer reactions in an expansion. The effect is close to zero after the start and quite steadily increases to the sixth quarter where it has its maximum with 0.03. Thereafter, it declines and gets close to zero again. Again, the shape of the reaction function of the linear model can be explained by the two states. The local minimum of the linear model around the third quarter is caused by the recession times and the local maximum around the sixth quarter can be explained by the expansion times.

It takes a few quarters until real capital formation reacts to a social security contribution shock in a recession. But after two quarters, capital formation declines up to -0.55 after five quarters. Hence, an unexpected decrease in the growth rate of social security contributions by one percentage point reduces real capital formation by 0.55 percent in the fifth quarter after the shock. After this minimum, the effect declines in size and gets close to zero. The reaction is similar to the one by private consumption but later in its timing, larger in its size and bumpier especially regarding the significance level. There seems to be no real reaction of capital formation after a social security contributions shock in an expansion. There is a weak decline until the second quarter and a subsequent increase up to the eighth quarter with a value of 0.3. The effect declines afterwards and is slightly, but statistically significantly, negative at the tenth and eleventh quarter. The LIRF finishes close to zero after twelve quarters. To a certain degree, this behavior is also similar to the reaction of private consumption. However, it is also larger in its size and later in its timing. This disentangling into the two states sheds some light to the linear model with its strange shape and significance levels that could not be interpreted easily.

For nominal wages, the effect of a social security contributions shock in a recession is negative right from the start. Again, this can partially be explained by the accounting method. The effect starts with -0.21 after the hit and declines to -0.38 after five quarters. That implies, that a sudden drop in the growth rate of social security contributions by one percentage point induces a decline of the nominal wage level by -0.38 at the fifth quarter after the hit. Then, the effect declines in its size until the eleventh quarter and increases again in the twelfth quarter. Only the effects around the minimum at the fifth quarter are statistically significant. Again, the effect is quite different when the unexpected reduction in social security contributions happens in a non-recessionary period. Then, the immediate effect is close to zero and starts to rise up to 0.10 after seven quarters and declines back to zero afterwards. None of the effects is statistically significant.

Given the accounting method of this gross wage scheme explained in Section 3.3.1, the net effect can be assumed positive in expansions. Since both estimated LIRFs have nearly the same quite flat behavior, they do not create a confusing shape of the LIRF in the linear model. But the effects of the expansionary state mitigate the effects of the recessionary state.

The effect of social security contributions on unemployment is miscellaneous. During recessions, a sudden drop in social security contributions also reduces unemployment. The effect has a first local minimum after one quarter with -0.02. An unexpected reduction of the growth rate of social security contributions by one percentage point reduces the unemployment rate by 0.02 percentage points. But this effects fades and even switches its sign such that there is a positive effect of 0.04 after five quarters. After that, there is a quite steady decline to -0.04 after eleven quarters. In expansions, the reaction is not much different but somewhat weaker within the first seven quarters. After seven quarters, there is a quite steady increase in the effect such that it closes at 0.04 after twelve quarters. In the expansionary state, the effects on unemployment are statistically insignificant. The diametrical behavior at the latter half of the series depending on the state explains the little kink in the linear model around the seventh quarter. The significant value in the beginning of the linear model might be induced by the additional statistical power that comes with more observations. This is the only counter-cyclical response I can observe in my estimations.

The negative responses during recessions are statistically insignificant but there is a theoretical explanation for the possibility of such a behavior of the economic agents. They might follow a forward-looking behavior and be aware of their current position in a recession. They could assume the government will finance the reduction in social security contributions with higher taxes (or social security contributions) in the future. As in the linear specification, the effects could be affected by the ZLB in the largest part of the time period.

3.2 reports the point estimators at three different horizons (four quarters, six quarters and eight quarters) for the components of the state-dependent model. Again, the numbers are provided to enhance comparability with other studies.

3.4 Conclusion

This paper shed some light on the effects of a change in social security contributions on macroeconomic variables. During a recession, the effects of a reduction in social security contributions are dominated by negative responses of the economy. I measure a small but significant decline in real GDP after four quarters. This might be caused by a forward-looking behavior of the economic agents that do not believe in the permanence of this reduction and prepare for an increase in the

Table 3.2: Response to social security contributions shock at different horizons (in %)

		4 quarters	6 quarters	8 quarters
Recession	Real GDP	-0.17*	-0.08	-0.03
	CPI	-0.02	0.03	0.00
	GDP deflator	-0.09	0.04**	0.05*
	Private consumption	-0.17*	-0.08	-0.10
	Capital formation	-0.31	-0.24	-0.01
	Wages	-0.35*	-0.37*	-0.32*
	Unemployment	0.03	0.00	-0.02
Expansion	Real GDP	0.03	0.06**	0.07*
	CPI	0.00	0.03*	0.02
	GDP deflator	-0.04	0.09***	0.03
	Private consumption	0.01	0.03	-0.02
	Capital formation	0.05	0.00	0.22
	Wages	0.01	0.05	0.07
	Unemployment	0.00	0.00	0.00

Notes: Level of significance: *** p<0.01, ** p<0.05, * p<0.1.

future. Moreover, prices increase for a transitory period of up to eight quarters. During a non-recessionary period, the effects found in this paper are quite different. My results indicate a small but significant increase in real GDP after approximately eight quarters. Again, I can measure a positive effect on the price level with a larger and more significant influence on the GDP deflator compared to the CPI.

There is some evidence in the microeconomic literature that the reaction by economic agents to a change in social security contributions is different than the reaction to a change in taxes. The economic agents might be aware that social security contributions are closer to an insurance contribution than to a tax. This paper gives some additional evidence from the macroeconomic perspective to this topic.

According to the findings of this paper and keeping in mind the existing literature on fiscal multipliers, a reduction of social security contributions as a stimulus in a recession cannot be seen as the best-fitting counteractive measure. A tax reduction - and even more additional government spending - seem to be better instruments to tackle an economic downturn. But is it desirable to cut social security contributions in an economic upswing? From a business cycle point of view, its pro-cyclical nature is a counter-argument. Independently from the position in the business cycle, reducing social security contributions is one of two possible ways in reducing tax wedges. Given the findings of this paper and considering the existing literature on tax multipliers, a reduction of the wedge should be executed via a reduction of the wage tax. It has more desirable effects.

Most of the time span examined in this paper is influenced by the ZLB. Whether this has an influence on the reaction to a change in social security contributions is left for future research.

Appendix A

Appendix - Survey-Based Structural Budget Balances

A.1 Additional Analysis

Revisions of the structural budget balance can be mainly attributed to two sources: the budget balance B and its correction for the business cycle, the cyclical component $\epsilon_{B,GAP}GAP$. In Section A.1.1, we show the revision size of the budget balance and compare it with the revision size of the cyclical component as shown in Figure 1.1. In Section A.1.1, we compare the relative revisions of the cyclical component with key figures of the System of National Accounts (SNA) to facilitate a classification of the revision size of the cyclical component.

A.1.1 Absolute Revisions

The revisions of the budget balance B and the cyclical component $\epsilon_{B,GAP}GAP$ are directly comparable although the former is in relation to GDP and the latter in relation to potential output. Equation (1.1) shows that a revision of the budget balance B of the amount z to $B+z$ influences the structural budget balance X to the same amount as a revision of $\epsilon_{B,GAP}GAP$ of the amount z to $\epsilon_{B,GAP}GAP+z$.[1] Figure A.1 shows the average size of the absolute value of an ex post revision of the budget balance between two subsequent vintages.[2] It contributes to the revisions of the structural budget balance to a somewhat smaller amount than the revisions of the cyclical component $\epsilon_{B,GAP}GAP$ as shown in Figure 1.1 in Section 1.3.1.

[1] $\epsilon_{B,GAP}GAP = \epsilon_{B,Y}Y$.

[2] The data for the budget balances in relation to nominal GDP was extracted from the IMF World Economic Outlook, resp. the OECD Economic Outlook for the vintages ranging from autumn 2003 to spring 2006.

However, the effect of these revisions of B on the structural budget balance X are also non-negligible. A possible reason for such large revisions at later stages might be changes in the System of National Accounts (SNA).

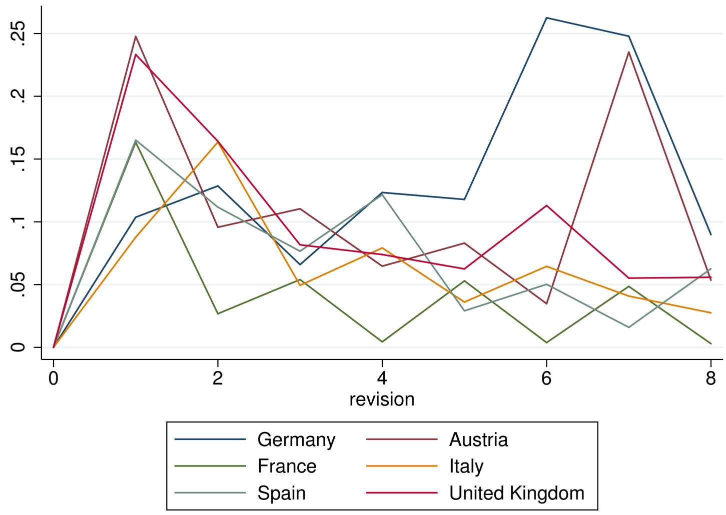


Figure A.1: Size of the ex post revisions of the budget balance (relative to nominal GDP) between two subsequent vintages (in percentage points of GDP)

Relative Revisions

Key figures of the SNA have a different unit than the cyclical component. Thus, we need to compare their relative revisions instead of the absolute ones. The relative absolute revision of a variable var is defined as $|(var_{t,i+1} - var_{t,i})/var_{t,i}|$. For the relative revision of GDP growth, we used the growth rates of GDP for the respective year: GDP_t/GDP_{t-1}. Figures A.2 to A.7 show that the relative revisions of the structural budget balances (production function and HP filter) are larger than the relative revisions of GDP (OECD and IMF), nominal GDP and the budget balance (relative to nominal GDP).[3] The HP filter series has a severe outlier.

[3]The data for the budget balances in relation to nominal GDP was extracted from the IMF World Economic Outlook, resp. the OECD Economic Outlook for the vintages ranging from autumn 2003 to spring 2006. GDP was extracted from vintages of the IMF World Economic Outlook and the OECD Economic Outlook. Nominal GDP was extracted from vintages of the IMF World Economic Outlook.

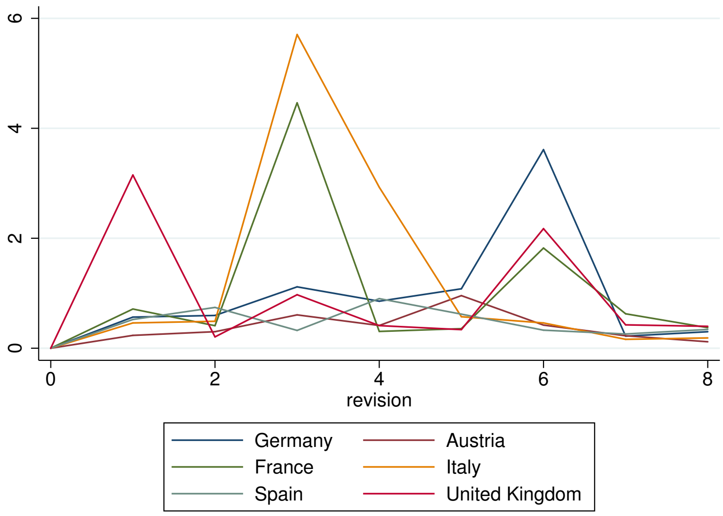


Figure A.2: Relative size of the ex post revisions of the structural budget balance between two subsequent vintages using the production function approach (in percentage to previous vintage)

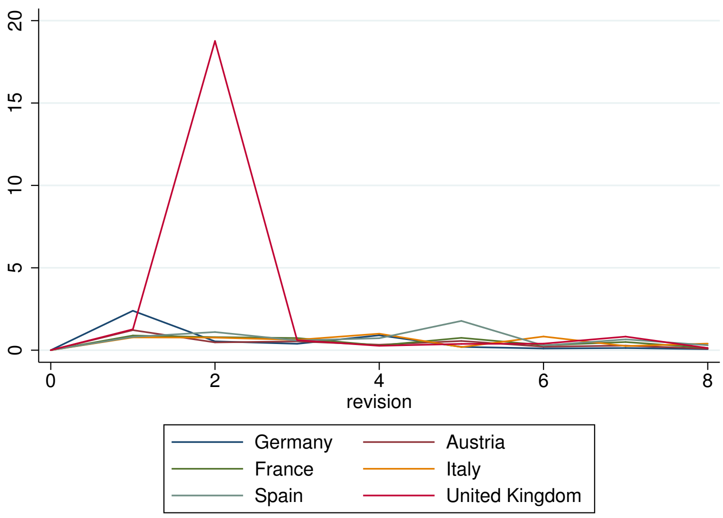


Figure A.3: Relative size of the ex post revisions of the structural budget balance between two subsequent vintages using an HP filter (in percentage to previous vintage)

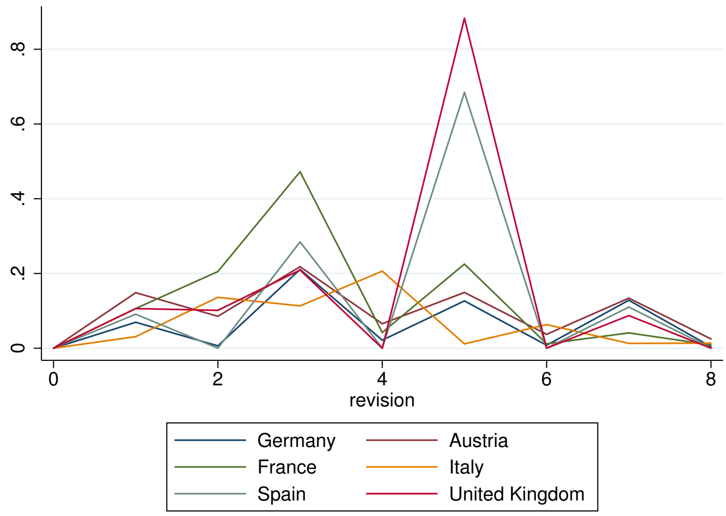


Figure A.4: Relative size of the ex post revisions of the growth of GDP (OECD) between two subsequent vintages (in percentage to previous vintage)

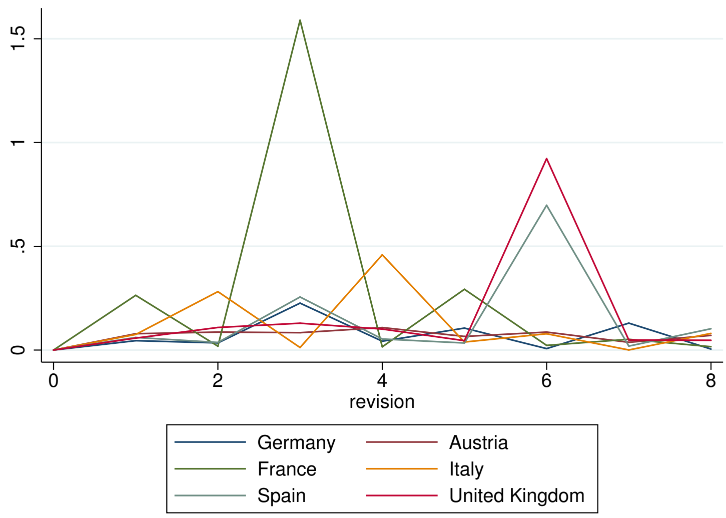


Figure A.5: Relative size of the ex post revisions of the growth of GDP (IMF) between two subsequent vintages (in percentage to previous vintage)

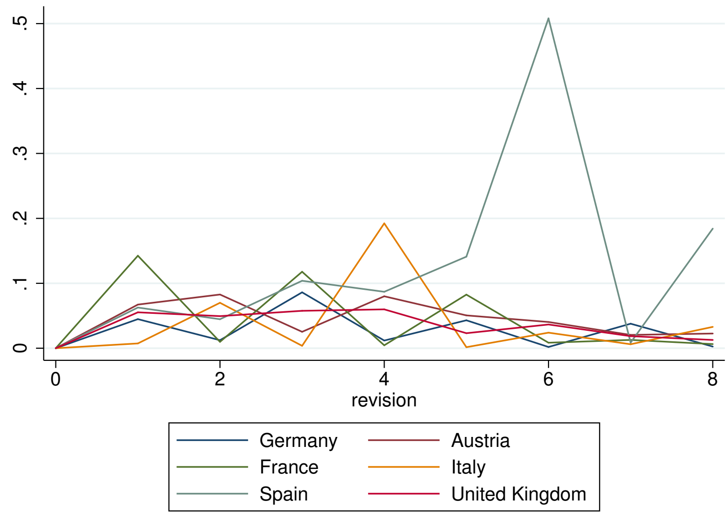


Figure A.6: Relative size of the ex post revisions of the growth of nominal GDP between two subsequent vintages (in percentage to previous vintage)

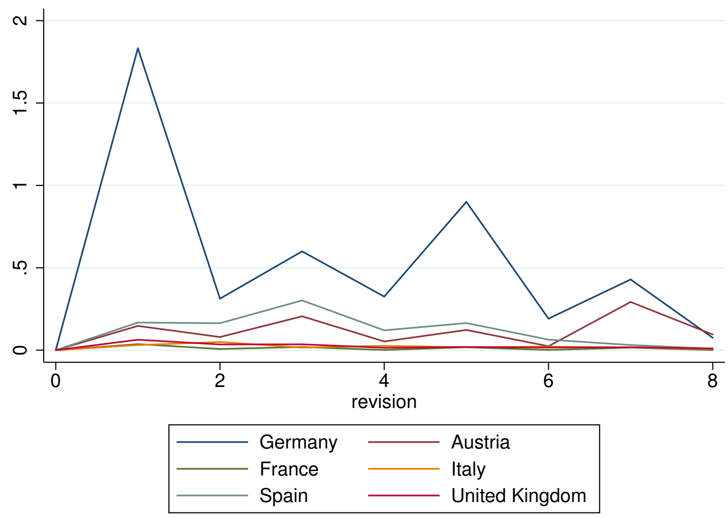


Figure A.7: Relative size of the ex post revisions of the budget balance (relative to nominal GDP) between two subsequent vintages (in percentage to previous vintage)

A.2 Data Adjustment

A.3 Backcasting Capacity Utilization in the Service Sector

For backcasting the series for capacity utilization in the service sector, we follow the approach suggested by Wohlrabe and Wollmershäuser (2017). They use a service sector confidence indicator that was already collected many years before capacity utilization in the European Commission's business surveys and which has a high correlation with the business cycle (see Table A.1 for the availability of the series in the Eurostat database).

Table A.1: First publication of confidence indicator in manufacturing and services

Sector	IT	DE	AT	FR	ES	UK
Manufacturing	1985q1	1985q1	1985q1	1985q1	1987q2	1985q1
Services	1998q1	1995q2	1996q4	1988q1	1996q4	1997q1

Source: Eurostat.

For backcasting capacity utilization we run the following regression

$$cu_{ser_{i,q}} = \alpha + \beta_i ci_{ser_{i,q}} + \epsilon_{i,q} \quad \text{(A.1)}$$

for every country i and quarter q, where cu_{ser} is the seasonally adjusted capacity utilization in the service sector, and ci_{ser} the seasonally adjusted services confidence indicator. In addition to country-specific analyses we also run a panel-regression.[4] The results for both, the panel and the country-specific regressions are summarized in Table A.2. All estimates for β_i are significant at the 1%-level and the R^2 range between 0.13 and 0.67.[5] For the sake of simplicity we use the panel estimate of β for backcasting capacity utilization with the values of the service sector confidence indicator. Figure A.8 shows the results of the backcast and compares the predicted series with the actual ones.

[4]Standard-errors are adjusted according to procedure proposed by Newey and West (1987) with a maximum lag length of four quarters. For every country, all available data points are used to extract a maximum of information. The maximum availability period spans from 1988q1 to 2019q1.

[5]A specification including a time trend gives a similar result, but shows no significance of the trend. Thus, we did not include a time trend in our baseline estimation.

Table A.2: Results for equation (A.1)

	Panel	AT	FR	DE	IT	ES	UK
α	87.870***	89.178***	91.004***	88.726***	87.471***	84.051***	87.525***
	(0.079)	(0.287)	(0.139)	(0.285)	(0.219)	(0.232)	(0.217)
β_i	0.094***	0.097***	0.083***	0.055***	0.109***	0.113***	0.061***
	(0.013)	(0.014)	(0.015)	(0.020)	(0.015)	(0.011)	(0.013)
Obs.	192	31	30	32	37	31	31
R^2	0.499	0.516	0.355	0.129	0.416	0.673	0.336

Notes: Standard errors in parentheses. Level of significance: *** p<0.01, ** p<0.05, * p<0.1. R^2: within R^2 for panel, adjusted R^2 for rest.

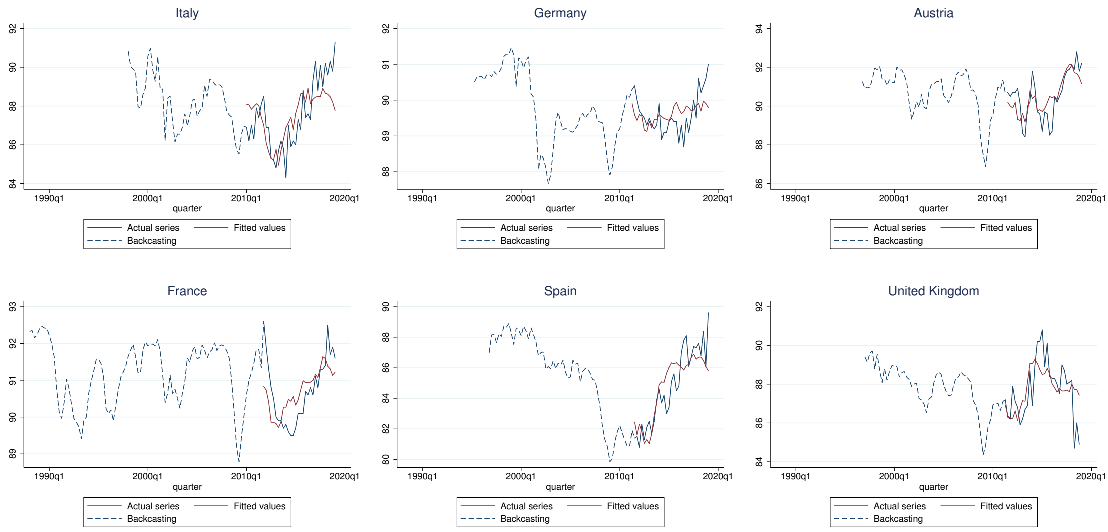


Figure A.8: Comparison of actual capacity utilization series in services with the fitted values (in percentage points)

As a robustness check, we re-estimate equation (A.1) for the manufacturing sector:

$$cu_{man_{i,q}} = \alpha + \beta_i ci_{man_{i,q}} + \epsilon_{i,q}. \tag{A.2}$$

Results are summarized in Table A.3. Figure A.9 compares the predicted series with the actual ones.

Table A.3: Results for equation (A.2)

	Panel	AT	FR	DE	IT	ES	UK
α	82.053***	86.218***	84.444***	85.618***	76.236***	79.617***	81.785***
	(0.120)	(0.124)	(0.199)	(0.179)	(0.145)	(0.194)	(0.165)
β_i	0.200***	0.205***	0.158***	0.245***	0.249***	0.213***	0.150***
	(0.020)	(0.023)	(0.025)	(0.021)	(0.025)	(0.026)	(0.019)
Obs.	745	93	113	137	137	128	137
R^2	0.496	0.634	0.364	0.641	0.540	0.422	0.427

Notes: Standard errors in parentheses. Level of significance: *** p<0.01, ** p<0.05, * p<0.1. R^2: within R^2 for panel, adjusted R^2 for rest.

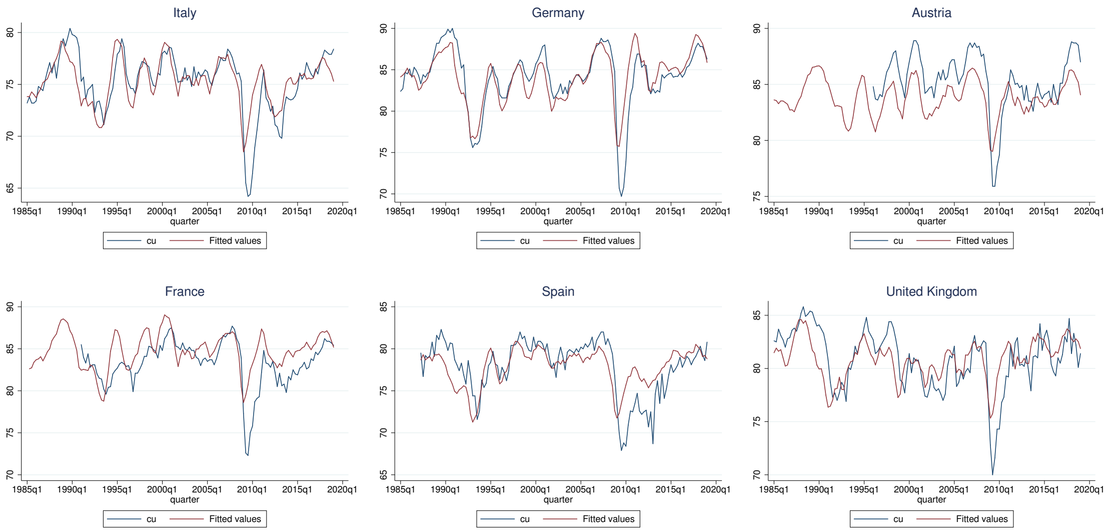


Figure A.9: Comparison of actual capacity utilization series in manufacturing with the fitted values (in percentage points)

A.4 Enlarging Capacity Utilization of Whole Economy via Manufacturing

If the series for capacity utilization in manufacturing in a country dates further back than the backcasted series for capacity utilization in services we use the manufacturing series to enlarge the overall capacity utilization. Specifically we assume that the first difference of the capacity utilization in manufacturing and in the overall economy is the same. The high correlation between the capacity utilization in manufacturing and the overall economy, both in levels (see Table A.4) and in first differences (see Table A.5) supports this approach.[6] This enlarged series is only used to calculate the mean of the capacity utilization series for the overall economy. We enlarge the series for Italy back to 1988q1, for Germany back to 1991q1, for Austria back to 1996q1, for Spain back to 1988q1 and for the United Kingdom back to 1988q1.

Table A.4: Correlation between capacity utilization in manufacturing and capacity utilization in whole economy

	Panel	IT	DE	AT	FR	ES	UK
Correlation	0.978	0.975	0.985	0.985	0.988	0.970	0.969
Observations	563	85	96	90	113	90	89

Table A.5: Correlation between first-differences of capacity utilization in manufacturing and of capacity utilization in whole economy

	Panel	IT	DE	AT	FR	ES	UK
Correlation	0.944	0.895	0.982	0.980	0.988	0.923	0.945
Observations	557	84	95	90	112	89	88

[6] For this evaluation, we used the original series for the service sector available for 2012-2018.

Appendix B

Appendix - Tax Revenue Forecast Errors

B.1 Details on the Applied Data

B.1.1 List of Taxes (resp. Tax Groups) Forecasted by the Working Party on Tax Revenue Estimates

- Wage tax *(Lohnsteuer)*
- Assessed income tax *(Veranlagte Einkommensteuer)*
- Not assessed income tax *(Nicht veranlagte Steuern vom Ertrag)*
- Withholding tax *(Abgeltungsteuer)*
- Corporate tax *(Körperschaftsteuer)*
- Sales taxes *(Steuern vom Umsatz)*
- Wealth tax *(Vermögensteuer)*
- Inheritance tax *(Erbschaftsteuer)*
- Land transfer tax *(Grunderwerbsteuer)*
- Racebetting and lottery tax *(Rennwett- und Lotteriesteuer)*
- Fireprevention tax *(Feuerschutzsteuer)*
- Beer tax *(Biersteuer)*
- Other State taxes *(sonstige Ländersteuern)*

- Business tax *(Gewerbesteuer)*
- Land tax A *(Grundsteuer A)*
- Land tax B *(Grundsteuer B)*
- Other community taxes *(Sonstige Gemeindesteuern)*
- Energy tax *(Energiesteuer)*
- Tobacco tax *(Tabaksteuer)*
- Alcohol tax *(Alkoholsteuer)*
- Alcopop tax *(Alkopopsteuer)*
- Sparkling wine tax *(Schaumweinsteuer)*
- Intermediate product tax *(Zwischenerzeugnissteuer)*
- Coffee tax *(Kaffeesteuer)*
- Insurance tax *(Versicherungsteuer)*
- Electricity tax *(Stromsteuer)*
- Motor vehicle tax *(Kraftfahrzeugsteuer)*
- Air traffic tax *(Luftverkehrsteuer)*
- Nuclear fuel tax *(Kernbrennstoffsteuer)*
- Solidarity surcharge *(Solidaritätszuschlag)*
- Lump-sum import duties *(Pauschalierte Einfuhrabgaben)*
- Other federal taxes *(Sonstige Bundessteuern)*

B.1.2 Data Sources

- Tax revenue forecasts: press releases of the Federal Ministry of Finance, biannual reports on the tax forecasts in the periodical *ifo Schnelldienst*
- Macroeconomic projections of the Federal Ministry for Economic Affairs and Energy: biannual reports on the tax forecasts by the the Federal Ministry of Finance in their monthly report, Working Party on Tax Revenue Estimates

- Projections of the Joint Economic Forecast: reports published in the periodical
ifo Schnelldienst

- Realized tax revenues: Federal Ministry of Finance

- Realized macroeconomic figures: Federal Statistical Office of Germany (vintage: Autumn 2020)

B.1.3 Data Availability

Table B.1: Sample length of the WPTRE's and the government's forecasts

Horizon	Taxes (sum & types)	Nominal GDP	Wages	Corporate Income	Private Consumption
$t + 0.5$	1992-2019	1992-2019	2003-2019	2006-2019	2011-2019
$t + 1$	1992-2019	1992-2019	2003-2019	2006-2019	2011-2019
$t + 1.5$	1992-2019	1992-2019	2003-2019	2006-2019	2012-2019
$t + 2$	1992-2019	1992-2019	2003-2019	2007-2019	2012-2019

Table B.2: Sample length of the JEF's forecasts

Horizon	Nominal GDP	Wages	Corporate Income	Private Consumption
$t + 0.5$	1992-2019	2001-2019	2001-2019	2001-2019
$t + 1$	1992-2019	2001-2019	2001-2019	2001-2019
$t + 1.5$	1992-2019	2002-2019	2002-2019	2002-2019
$t + 2$	1992-2019	2002-2019	2002-2019	2002-2019

B.1.4 Correlation Across Forecast Errors

Table B.3: Correlation Across FE^B and FE^ε

	Tax Sum	Wage Tax	Income Tax	Business Tax	Corpor. Tax	Energy Tax	Sales Taxes
Corr.	0.24	-0.10	0.25	0.15	0.07	-0.13	-0.53
Obs.	112	68	55	55	55	112	34

B.2 Absolute Explanatory Power

Table B.1: Squared Semi-Partial Correlation Coefficients – WPTRE

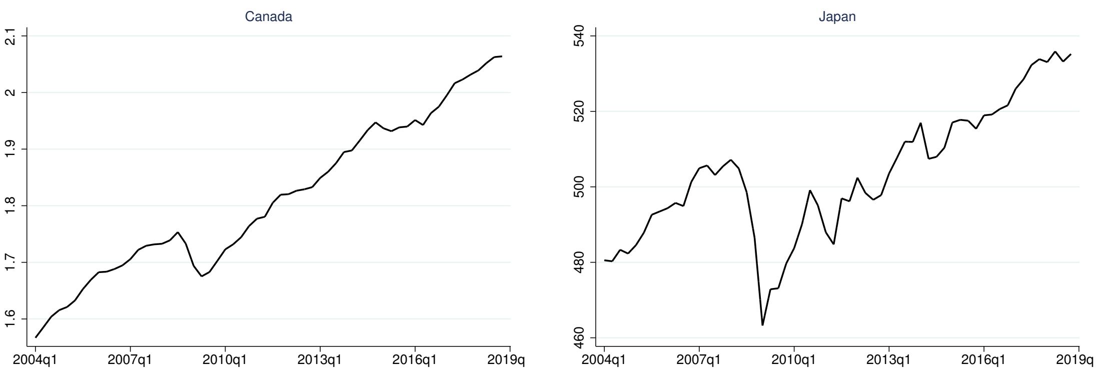
	Tax Sum	Wage Tax	Income Tax	Business Tax	Corpor. Tax	Energy Tax	Sales Taxes
FE^B	0.43	0.65	0.24	0.30	0.39	0.03	0.43
FE^ε	0.19	0.05	0.00	0.00	0.01	0.43	0.80

Appendix C

Appendix - The State-Dependent Effects of Social Security Contributions on the Macroeconomy

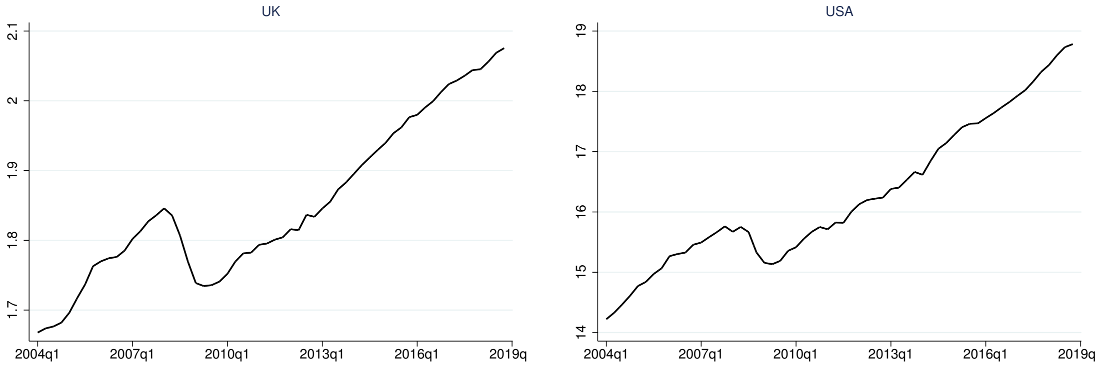
C.1 Additional Descriptive Graphs

Figure C.1: Real gross domestic product



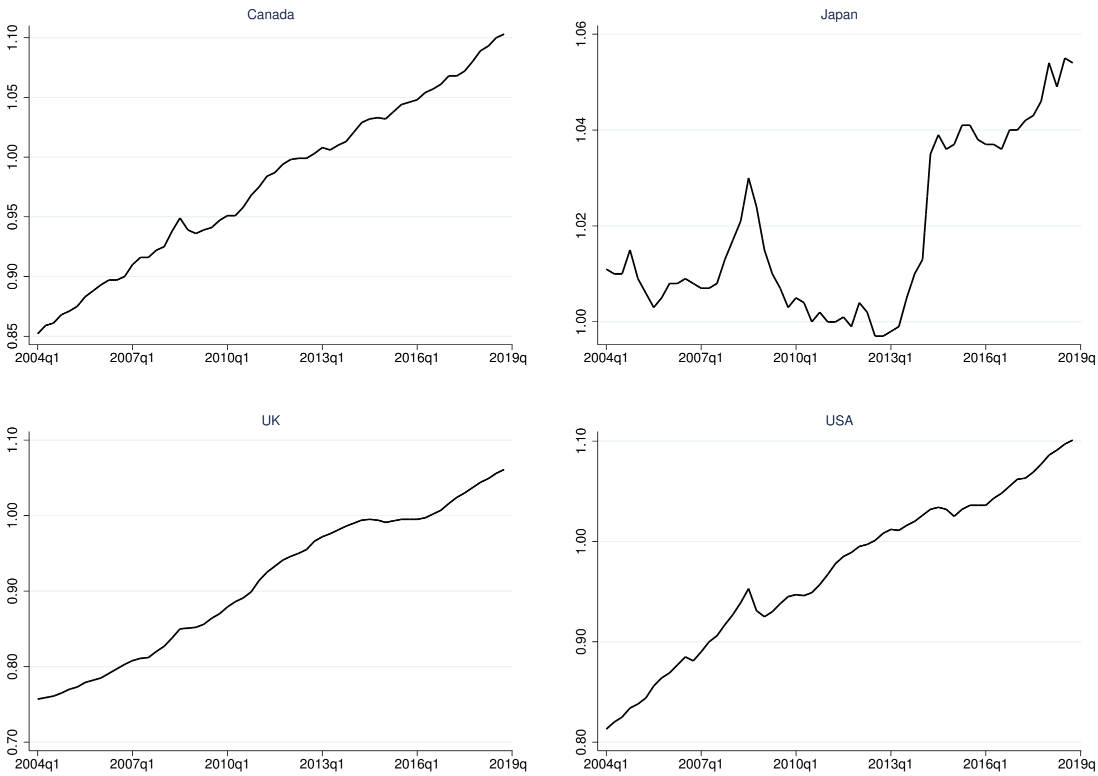
Notes: The figure shows real GDP (in trn. of the national currency) in the respective countries between 2004q1 and 2018q4.

Figure C.2: Real gross domestic product (cont.)



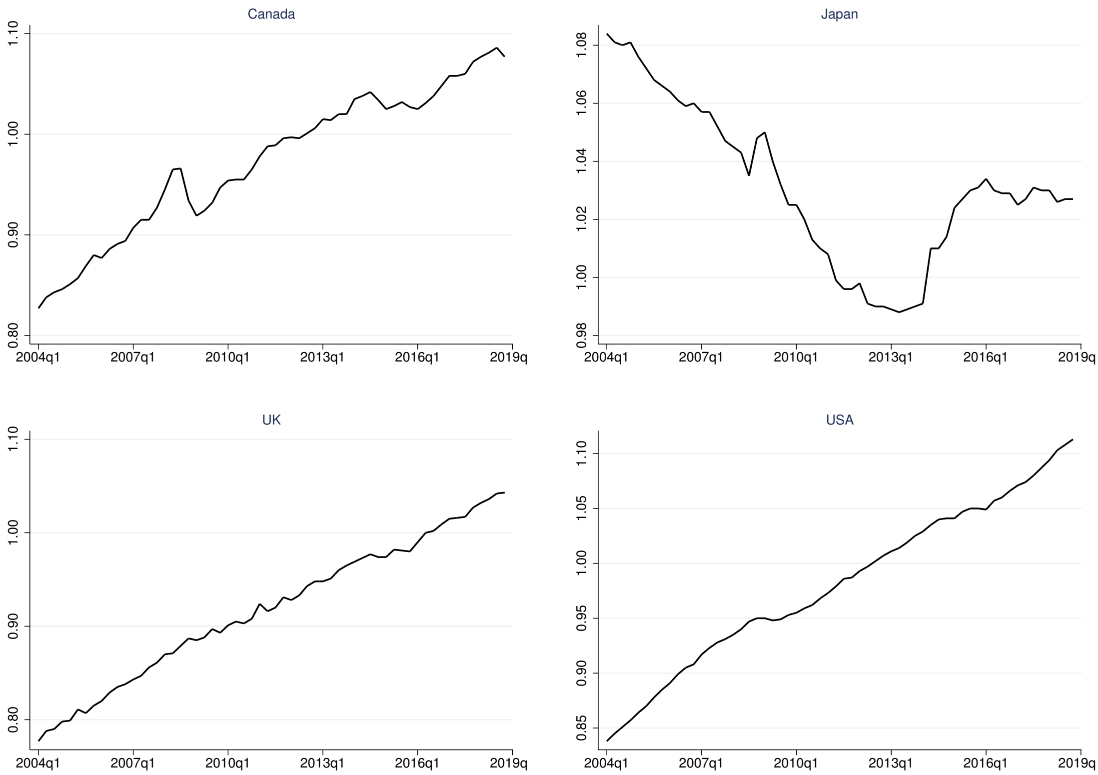
Notes: The figure shows real GDP (in trn. of the national currency) in the respective countries between 2004q1 and 2018q4.

Figure C.3: Consumer price index



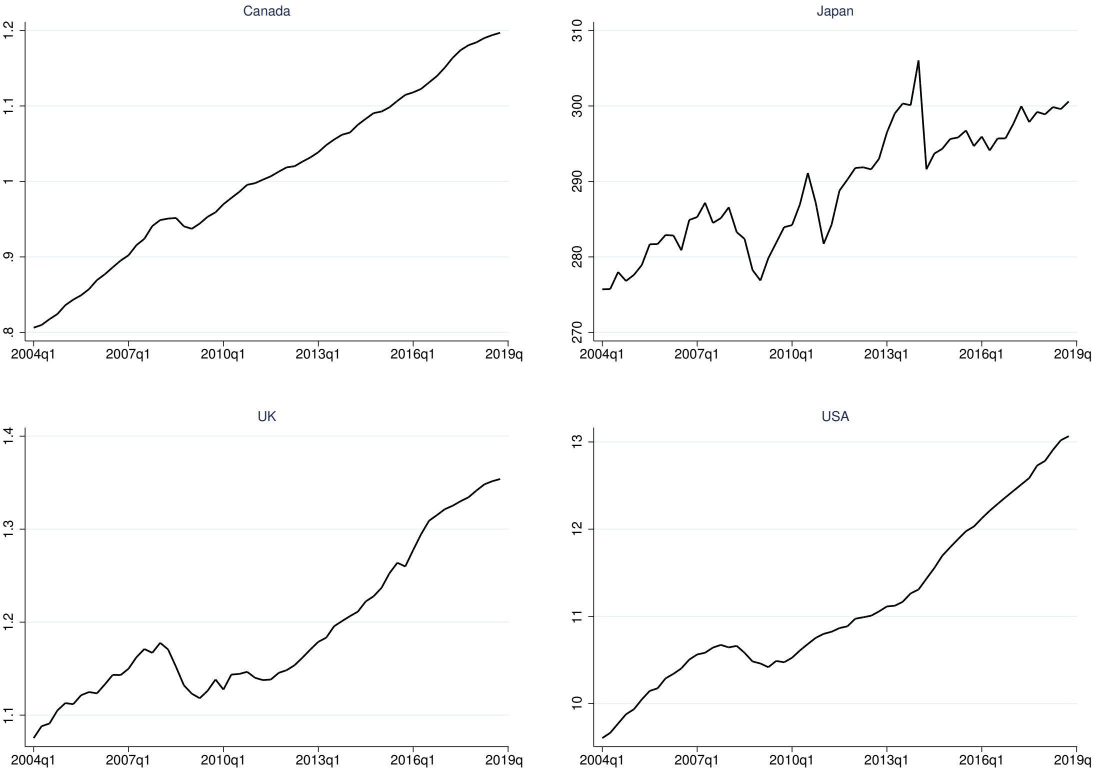
Notes: The figure shows the CPI in the respective countries between 2004q1 and 2018q4.

Figure C.4: Gross domestic product deflator



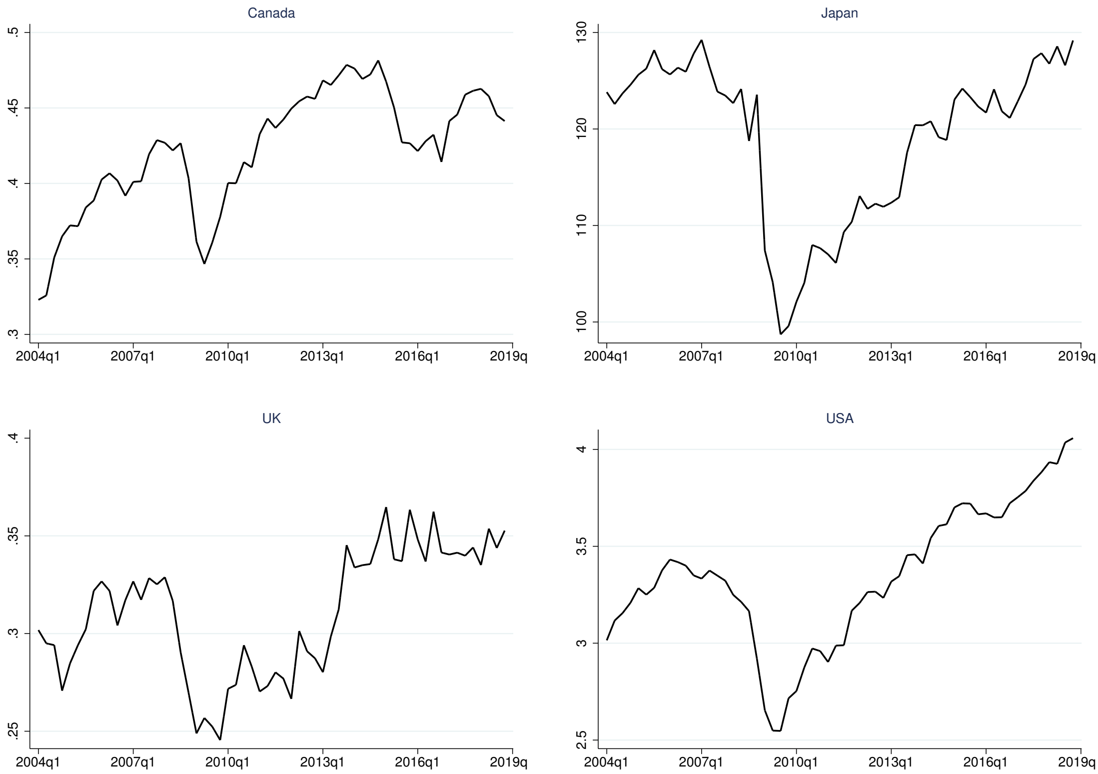
Notes: The figure shows the GDP deflator in the respective countries between 2004q1 and 2018q4.

Figure C.5: Real private consumption



Notes: The figure shows real private consumption (in trn. of the national currency) in the respective countries between 2004q1 and 2018q4.

Figure C.6: Real capital formation

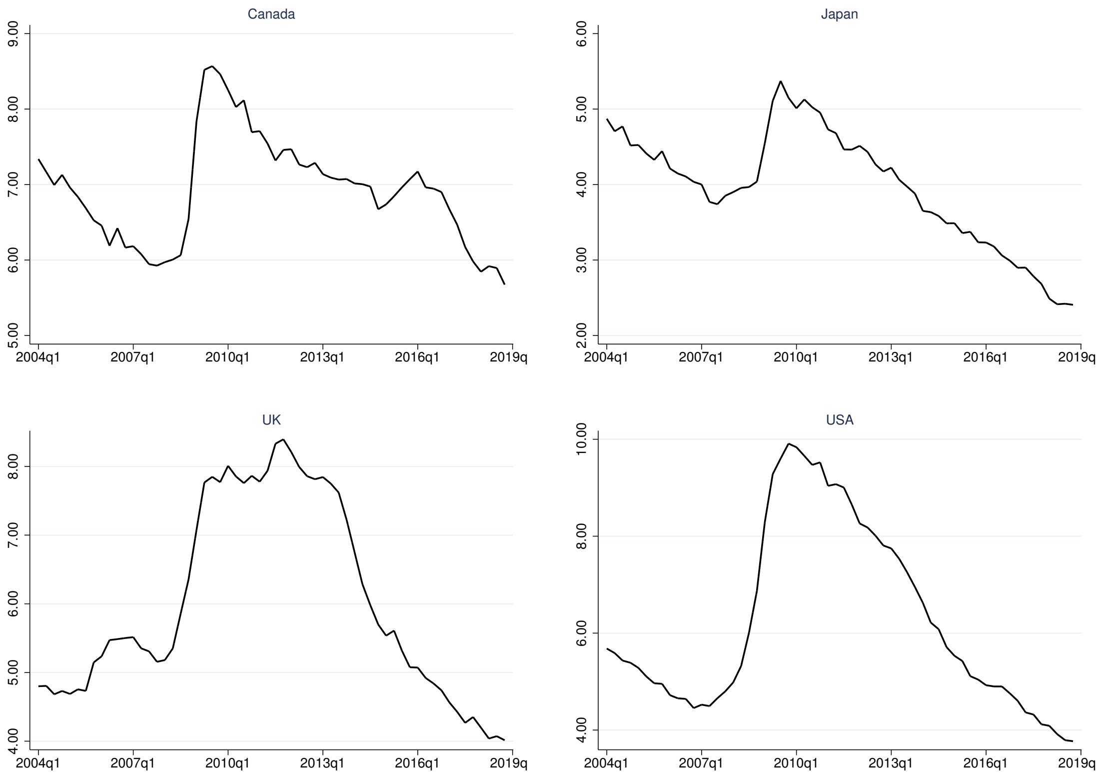


Notes: The figure shows real capital formation (in trn. of the national currency) in the respective countries between 2004q1 and 2018q4.

Figure C.7: Nominal wages

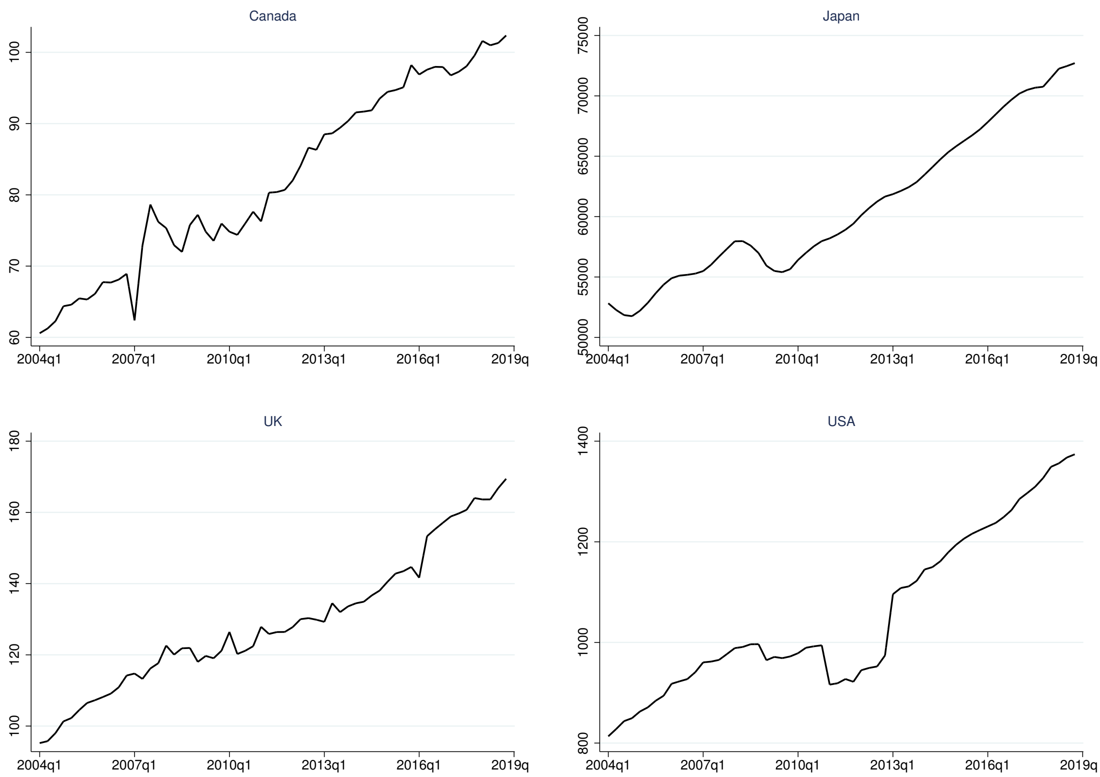
Notes: The figure shows the sum of gross nominal wages (in trn. of the national currency) in the respective countries between 2004q1 and 2018q4.

Figure C.8: Unemployment



Notes: The figure shows the unemployment rate (in %) in the respective countries between 2004q1 and 2018q4.

Figure C.9: Nominal social contributions



Notes: The figure shows the realized nominal social security contributions (in bln. of the national currency) in the respective countries between 2004q1 and 2018q4.

Bibliography

Ademmer, E. and Dreher, F. (2016). Constraining Political Budget Cycles: Media Strength and Fiscal Institutions in the Enlarged EU. *Journal of Common Market Studies*, **54** (3), 508–524.

Ademmer, M. and Boysen-Hogrefe, J. (2022). The Impact of Forecast Errors on Fiscal Planning and Debt Accumulation, Kiel Working Papers 2123.

Alpanda, S., Granziera, E. and Zubairy, S. (2021). State Dependence of Monetary Policy Across Business, Credit and Interest Rate Cycles. *European Economic Review*, **140**.

Asimakopoulos, S., Paredes, J. and Warmedinger, T. (2020). Real-Time Fiscal Forecasting Using Mixed-Frequency Data. *Scandinavian Journal of Economics*, **122** (1), 369–390.

Auerbach, A. (1999). On the performance and use of government revenue forecasts. *National Tax Journal*, **52** (4), 765–782.

— and Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, **4** (2), 1–27.

— and — (2013). Fiscal Multipliers in Recession and Expansion. In A. Alesina and F. Giavazzi (eds.), *Fiscal Policy after the Financial Crisis*, Chicago: University of Chicago Press, pp. 63–98.

Batchelor, R. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, **23** (2), 189–203.

Beetsma, R., Bluhm, B., Giuliodori, M. and Wierts, P. (2013). From budgetary forecasts to ex post fiscal data: Exploring the evolution of fiscal forecast errors in the European Union. *Contemporary Economic Policy*, **31** (4), 795–813.

—, Giuliodori, M. and Wierts, P. (2009). Planning to cheat: EU fiscal policy in real time. *Economic Policy*, **24** (60), 753–804.

Bergmann, U. M., Hutchison, M. M. and Jensen, S. E. H. (2016). Government Efficiency, Institutions, and the Effects of Fiscal Consolidation on Public Debt. *European Journal of Political Economy*, **44**, 1–19.

Bischoff, I. and Gohout, W. (2010). The political economy of tax projections. *International Tax and Public Finance*, **17** (2), 133–150.

Blanchard, O. and Perotti, R. (2002). An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *Quarterly Journal of Economics*, **117** (4), 1329–1368.

Breuer, C. (2015). On the Rationality of Medium-Term Tax Revenue Forecasts: Evidence from Germany. *Journal of Economics and Statistics*, **235** (1), 313–340.

Bundesbank (2014). *On the reliability of international organisations' estimates of the output gap*. Monthly Report 04-2014, Bundesbank.

Buti, M. and Carnot, N. (2012). The EMU Debt Crisis: Early Lessons and Reforms. *Journal of Common Market Studies*, **50** (6), 899–911.

—, —, Hristov, A., McMorrow, K., Roeger, W. and Vandermeulen, V. (2019). Potential output and EU fiscal surveillance . `https://voxeu.org/article/potential-output-and-eu-fiscal-surveillance`, [Online, accessed 25-September-2019].

Büttner, T. and Kauder, B. (eds.) (2008). *Methoden der Steuerschätzung im internationalen Vergleich*. ifo Forschungsberichte No. 44, ifo Institute, Munich.

— and — (2010). Revenue Forecasting Practices: Differences across Countries and Consequences for Forecasting Performance. *Fiscal Studies*, **31** (3), 313–340.

— and — (2015). Political Biases despite Extermal Expert Participation? An Empirical Analysis of Tax Revenue Forecasts in Germany. *Public Choice*, **164** (3/4), 287–307.

Caggiano, G., Castelnuovo, E., Colombo, V. and Nodari, G. (2015). Estimating Fiscal Multipliers: News from a Non-Linear World. *Economic Journal*, **125** (5), 746–776.

Capek, J. and Crespo Cuaresma, J. (2020). We Just Estimated Twenty Million Fiscal Multipliers. *Oxford Bulletin of Economics and Statistics*, **82** (3), 483–502.

—, —, Holler, J. and Schuster, P. (2021). Fiscal Multipliers in a Small Open Economy: the Case of Austria. *Oxford Economic Papers*, pp. 1–19.

Caselli, F. and Reynaud, J. (2020). Do fiscal Rules Cause Better Fiscal Balances? A new Instrumental Variable Strategy. *European Journal of Political Economy*, **63**.

Cayen, J.-P. and van Norden, S. (2005). The Reliability of Canadian Output-Gap Estimates. *North American Journal of Economics and Finance*, **16**, 373–393.

Chahrour, R., Schmitt-Grohe, S. and Uribe, M. (2012). A Model-Based Evaluation of the Debate on the Size of the Tax Multiplier. *American Economic Journal: Economic Policy*, **4** (2), 28–45.

Christiano, L., Eichenbaum, M. and Rebelo, S. (2011). When is the Government Spending Multiplier Large. *Journal of Political Economy*, **119** (1), 78–121.

Christofzik, D., Fuest, A. and Jessen, R. (2022). Macroeconomic Effects of the Anticipation and Implementation of Tax Changes in Germany: Evidence from a Narrative Account. *Economica*, **89** (1), 62–81.

Cimadomo, J. (2016). Real-time data and fiscal policy analysis: A survey of the literature. *Journal of Economic Surveys*, **30** (2), 302–326.

DeJong, J. F. M. and Gilbert, N. D. (2020). Fiscal discipline in EMU? Testing the effectiveness of the Excessive Deficit Procedure. *European Journal of Political Economy*, **61**.

deResende, C. (2014). *An Assessment of IMF Medium-Term Forecasts of GDP Growth*. Background Paper -February 2012, Independent Evaluation Office of the IMF.

EC (2018). *Vade Mecum on the Stability and Growth Pact*. European Economy Institutional Paper - March 2018 075, European Commission.

ECB (2007). Prices and Costs - Capacity Utilisation in the Euro Area: Measures and Implications. `https://www.ecb.europa.eu/pub/pdf/other/mb200710_focus05.en.pdf?be6b47dfc034153534fd0e4d34cd7581`, [Online; accessed 09-July-2019].

ECOFIN (2017). Specifications on the Implementation of the Stability and Growth Pact and Guidelines on the Format and Content of Stability and Convergence Programmes. `http://data.consilium.europa.eu/doc/document/ST-9344-2017-INIT/en/pdf`, [Online, accessed 12-July-2019].

EU (2012). Treaty on Stability, Coordination and Governance in the Economic and Monetary Union. https://www.consilium.europa.eu/media/20399/st00tscg26_en12.pdf, [Online, accessed 07-October-2020].

Evans, K. (1969). Reconstruction and Estimation of the Balanced Budget Multiplier. *Review of Economics and Statistics*, **51** (1), 14–25.

Favero, C. and Giavazzi, F. (2012). Measuring Tax Multipliers: The Narrative Method in Fiscal VARs. *American Economic Journal: Economic Policy*, **4** (2), 69–94.

Fedelino, A., Ivanova, A. and Horton, M. (2009). *Computing Cyclically Adjusted Balances and Automatic Stabilizers*. Technical notes and manuals, IMF.

Fioramanti, M. and Waldmann, R. (2016). The Stability and Growth Pact: Econometrics and its consequences for human beings . https://voxeu.org/article/econometrics-and-its-consequences-human-beings, [Online, accessed 12-July-2019].

Frankel, J. and Schreger, J. (2016). Bias in Official Forecasts: Can Private Forecasts Help?, NBER Working Paper No. 22349.

Gebhardt, H. (2001). Methoden, Probleme und Ergebnisse der Steuerschätzungen. *RWI-Mitteilungen*, **52** (2), 127–147.

Gechert, S. (2015). What Fiscal Policy is Most Effective? A Meta-Regression Analysis. *Oxford Economic Papers*, **67** (3), 553–580.

—, Paetz, C. and Villanueava, P. (2021). The Macroeconomic Effects of Social Security Contributions and Benefits. *Journal of Monetary Economics*, **117**, 571–584.

Girouard, N. and André, C. (2005). *Measuring Cyclically-Adjusted Budget Balances for OECD Countries*. OECD Economics Department Working Papers No. 434, OECD.

Göttert, M. and Lehmann, R. (2021). *Tax Revenue Forecast Errors: Wrong Predictions of the Tax Base or the Elasticity?* CESifo Working Paper No. 9148, CESifo.

— and Wollmershäuser, T. (2021). *Survey-Based Structural Budget Balances*. CESifo Working Paper No. 8911, CESifo.

Graff, M. and Sturm, J.-E. (2012). The Information Content of Capacity Utilization Rates for Output Gap Estimates. *CESifo Economic Studies*, **58** (12), 220–251.

Havik, K., McMorrow, K., Orlandi, F., Planas, C., Raciborski, R., Roeger, W., Rossi, A., Thum-Thysen, A. and Vandermeulen, V. (2014). *The Production Function Methodology for Calculating Potential Growth Rates and Output Gaps*. European Economy Economic Paper 535-November 2014, European Commission.

Heinemann, F. (2006). Planning or Propaganda? An Evaluation of Germany‘s Medium-term Budgetary Planning. *FinanzArchiv: Public Finance Analysis*, **63** (4), 551–578.

Hodrick, R. J. and Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, **29**, 1–16.

Holden, D. and Peel, K. A. (1990a). On Testing for Unbiasedness and Efficiency of Forecasts. *The Manchester School*, **58**, 120–127.

Holden, K. and Peel, D. A. (1990b). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, **58** (2), 120–127.

Hulej, M. and Grabek, G. (2015). *Output Gap Measure Based on Survey Data*. NBP Working Paper 200, Narodowy Bank Polski.

IMF (1993). *Annex I, Structural Budget Indicators for the Major Industrial Countries*. World Economic Outlook 1993/2, IMF.

Jochimsen, B. and Lehmann, R. (2017). On the political economy of national tax revenue forecasts: evidence from OECD countries. *Public Choice*, **170** (3/4), 211–230.

Jonung, L. and Larch, M. (2006). Improving fiscal policy in the EU: The case for independent forecasts. *Economic Policy*, **21** (47), 491–534.

Jorda, O. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, **95** (1), 161–182.

Kaiser, R. and Maravall, A. (2001). *Measuring Business Cycles in Economic Time Series*. Lecture Notes in Statistics - Springer 154.

Kauder, B., Potrafke, N. and Schinke, C. (2017). Manipulating Fiscal Forecasts: Evidence from the German States. *FinanzArchiv: Public Finance Analysis*, **73** (2), 213–236.

Keane, M. P. and Runkle, D. E. (1989). Are Economic Forecasts Rational? *Federal Reserve Bank of Minneapolis Quarterly Review*, **13** (2), 26–33.

— and — (1990). Testing the Rationality of Price Forecasts: New Evidence from Panel Data. *American Economic Review*, **80** (4), 714–735.

Kempkes, G. (2014). Cyclical Adjustment in Fiscal Rules: Some Evidence on Real-Time Bias for EU-15 Countries. *Public Finance Analysis*, **70** (2), 278–315.

Leal, T., Pérez, J. J., Tujula, M. and Vidal, J. P. (2008). Fiscal foreacsting: Lessons from the literature and challenges. *Fiscal Studies*, **29** (3), 347–386.

Lehmann, R. and Wollmershäuser, T. (2020). The Macroeconomic Projections of the German Government: A Comparison to an Independent Forecasting Institution. *German Economic Review*, **21** (2), 235–270.

Lenain, P. (2002). What is the Track Record of OECD Economic Projections? Paris: OECD.

Maravall, A. and del Río, A. (2001). *Time Aggregation and the Hodrick-Prescott Filter*. Documento de Trabajo 0108, Banco de España.

Marcellino, M. and Musso, A. (2011). The Reliability of Real-Time Estimates of the Euro Area Output Gap. *Economic Modelling*, **28**, 1842–1856.

McMorrow, K., Roeger, W., Vandermeulen, V. and Havik, K. (2015). *An assessment of the relative quality of the Output Gap estimates produced by the EU's Production Function Methodology*. Discussion Paper-December 2015 020, European Commission.

Mincer, J. A. and Zarnowitz, V. (1969). The Evaluation of Economic Forecasts. In J. A. Mincer (ed.), *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, Studies in Business Cycles, *1*, National Bureau of Economic Research, pp. 3–46.

Mise, E., Kim, T.-H. and Newbold, P. (2005). On Suboptimaility of the Hodrick-Prescott filter at time series endpoints. *Journal of Macroeconomics*, **27** (1), 53–67.

Mourre, G., Poissonier, A. and Lausegger, M. (2019). *The Semi-Elasticities Underlying the Cyclically-Adjusted Budget Balance: An Update and Further Analysis*. Discussion Paper 098-May 2019, European Commission.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**, 703–708.

Nordhaus, W. D. (1987). Forecasting Efficiency: Concepts and Applications. *Review of Economics and Statistics*, **69** (4), 667–674.

Nyman, C. (2010). *An Indicator of Resource Utilisation*. Economic Commentaires 04/2010, Sveriges Riksbank.

OECD (2012). *Chapter 4, Medium and long-term scenarios for global growth and imbalances*. Economic Outlook 2012/1, OECD.

Orphanides, A., Porter, R. D., Reifschneider, D., Tetlow, R. and Finan, F. (2000). Errors in the measurement of the output gap and the design of monetary policy. *Journal of Economics and Business*, **52** (1-2), 117–141.

— and van Norden, S. (2002). The Unreliability of Output-Gap Estimates in Real Time. *Review of Economics and Statistics*, **84**, 569–583.

Owyang, M., Ramey, V. and S., Z. (2013). Are Government Spending Multipliers Greater During Periods of Slack? Evidence from Twentieth Century Hstorical Data. *American Economic Review*, **103** (3), 129–134.

Plagborg-Moller, M. and Wolf, C. (2020). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, **89** (2), 955–980.

Price, R., Dang, T. and Guillemette, Y. (2014). *New Tax and Expenditure Elasticty Estimates for EU Budget Surveillance*. OECD Economics Department Working Papers 1174, OECD.

Ramey, V. (2012). Comment on Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment. *NBER Macroeconomics Annual 2012*, **27**, 147–153.

— and Zubairy, S. (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data. *Journal of Political Economy*, **126** (2), 850–901.

Roeger, W., McMorrow, K., Hristov, A. and Vandermeulen, V. (2019). *Output Gaps and Cyclical Indicators*. Discussion Paper 104-July 2019, European Commission.

ROMER, C. and ROMER, D. (2010). The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review*, **100** (3), 763–801.

SAEZ, E., MATSAGANIS, M. and TSAKLOGLOU, P. (2012). Earninga Determination and Taxes: Evidence from a Cohort-Based Payroll Tax Reform in Greece. *Quarterly Journal of Economics*, **127** (1), 493–533.

SILVA, D., MONTEIRO, C., LEAL, J. and GUTIÉRREZ, E. (2016). *A Survey-Based measure of Output Gap for Portugal*. Artigos da autoria do GPEARI 06/2016, Gabinete de Planeamento, Estratéia, Avaliação e Relações Internacionais, Ministério das Finançias.

SIMS, E. and WOLFF, J. (2018). The State-Dependent Effects of Tax Shocks. *European Economic Review*, **107** (C), 57–85.

STOCK, J. and WATSON, M. (2007). Why has US Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, **39** (1), 3–33.

STRAUCH, R., HALLERBERG, M. and VON HAGEN, J. (2004). Budgetary forecasts in Europe – The track record of Stability and Convergence Programmes, ECB Working Paper Series No. 307.

SZÖRFI (2015). A survey-based measure of slack. `http://ec.europa.eu/economy_finance/events/2015/20150928_workshop/pdf/paper_bela_szorfi.pdf`, [Online; accessed 09-July-2019].

TOOZE, A. (2019). Output gap nonsense. `https://www.socialeurope.eu/output-gap-nonsense`, [Online, accessed 25-September-2019].

VOGEL, L. (2007). How do the OECD Growth Projections for the G7 Econonomies Perform? A Post-Mortem, OECD Working Paper no. 573. Paris: OECD.

WOHLRABE, K. and WOLLMERSHÄUSER, T. (2017). Zur Konstruktion einer gesamtwirtschaftlichen ifo Kapazitätsauslastung. *ifo Schnelldienst*, **70** (15), 26–30.