Fiscal Policy and Recessions: Effects on Economic Growth and Potential Output

Inauguraldissertation

zur Erlangung des akademischen Grades
Doctor Rerum Politicarum

an der Fakultät für Wirtschafts- und Sozialwissenschaften der
Ruprecht-Karls-Universität Heidelberg

vorgelegt von

Christopher Paul Zuber

Heidelberg, May 2022

Acknowledgements

In preparing this thesis, I received support from many people to whom I am most grateful. First and foremost, I wish to express my sincere gratitude to my primary supervisor, Jonas Dovern. Thank you for your continuous guidance and your dedicated support over all the years. I appreciated your honest feedback at all stages of this thesis, and I am very happy that our two papers are published. Second, I am deeply grateful to Zeno Enders who adopted me when Jonas left for Nuremberg. I always enjoyed our collaboration in teaching, and I am very thankful for your valuable career advice.

Special thanks go to my co-author David Vespermann. It was a pleasure to write our paper together, and I will miss our discussions and shared office. I would also like to thank Switgard Feuerstein for the opportunity to teach mathematics, and Alexander Glas and Onno Kleen for the great time together.

Conducting research at the Alfred-Weber-Institute for Economics was a privilege, not least because of the stimulating research environment in the Macro & Econometrics group. I benefited a lot from interactions with Sarah Arndt, Christian Conrad, Thomas Eife, Matthias Hartmann, Fabian Krüger, Joep Lustenhouwer, Lora Pavlova, Jörg Rieger, Sebastian Rüth, and Michael Stollenwerk. I also thank all other people at the Alfred-Weber-Institute for making it a great place to work.

I am thankful to friends and family who distracted me and thereby helped me to approach my research with a fresh mind. I thank Nils, especially for wonderful journeys, Langensteins where I was always welcome, the inhabitants of Wiesberg who remind me where I am from, Alex for enjoyable conversations, Ihno and Rebecca for inspiring climbing sessions, and Febian and Lukas for great years while sharing a flat.

This thesis would not have been possible without the unconditional support of four outstanding people. Hendrik, thank you very much for being my counterpart in countless discussions and for always giving excellent advice. Pia, your love and understanding was fundamental, especially at times of frustration. Time without you would have been so much less exciting, and I am looking forward to our next destination. Finally, I am greatly indebted to my parents. Thank you that you always believed in me. I owe you who I am today. This thesis is dedicated to you.

Contents

List of Figures

List of Tables

1

Introduction

Business cycles are characterized by alternating periods of recessions and expansions. In these periods, the output gap, which measures the deviation of output from its sustainable long-run potential, often systematically differs. During expansions, demand rises and firms increase production up to a level at which output exceeds the long-run sustainable production capacity. The output gap is positive, and high demand for production inputs like labor and capital can lead to inflationary pressure. During recessions, general economic activity and the employment level decline. In turn, the income of households drops, and increased uncertainty about the future can lead to delays of larger purchases or investments. The output gap is negative indicating slack in the economy.

The degree of slack in the economy, measured by the output gap, and the sustainable economic long-run trend, measured by potential output, convey important information for fiscal and monetary policy. Policy makers therefore monitor estimates of the output gap and potential output closely. Central banks, for example, assess the output gap to decide on the interest rate level. A positive output gap indicates that the economy overheats, and the central bank may raise interest rates to dampen inflationary pressure. Fiscal authorities are concerned whether current and future government spending paths are sustainable. Prolonged phases of negative output gaps indicate lower employment levels and reduced tax revenues. Similarly, downward revisions of potential output signal lower tax income paths in the future. In response, the government needs to balance budgets or guarantee that it will be able to service the debt in the future which it accumulates by running a deficit today. In addition, the government may consider countercyclical fiscal policies to stabilize the economy and judge on the economy's position in the business cycle based on the output gap.

The scenarios show that reliable estimates of the output gap and potential output are important resources for authorities to decide on fiscal and monetary policies. However, neither potential output nor the output gap is observable. The unreliability of output gap estimates is well documented (e.g., Orphanides and van Norden, 2002; Orphanides, 2003). Less is known about revisions of potential output levels after recessions. In addition, the empirical evidence on how the effect of fiscal stimulus depends on economic conditions is mixed. In this thesis, I therefore study the effects of fiscal policy and recessions on economic growth and potential output. Chapters 2 and 3 document downward revisions of potential output estimates after recessions. In Chapter 4, the focus is on how reliable estimates of the output gap can be obtained. Chapter 5 shows that the effect of fiscal policy depends on the state of the economy. Each chapter thereby represents a self-contained research article.

Chapter 2, co-authored by Jonas Dovern, investigates expert revisions of potential output estimates for 95 recessionary episodes in 27 developed countries. Although potential output should be independent of cyclical fluctuations, Benati (2012) and Ball (2014), for example, show that potential output estimates are revised downwards for many countries after the Great Recession. We outline four potential explanations for the revisions of potential output estimates. First, supply shocks may lower potential output as previous expectations about the long-run output path decrease permanently. Second, temporary demand shocks can lead to hysteresis, for example, in the labor market. Third, revisions of potential output estimates may solely reflect the correction of previous measurement error. Fourth, revisions may happen because of reverse causality. For example, Blanchard et al. (2015, 2017) suggest that pessimistic income expectations lower demand which eventually leads into a recession.

Using real-time data from the Organisation for Economic Co-operation and Development, we analyze how likely each of the four explanations is. We show that downward revisions of potential output estimates are substantial, permanent, and mostly driven by supply shocks. In contrast, estimates of potential output do not significantly react to demand shocks. Revisions are also partly caused by mismeasurement of potential output before recessions. In particular, we show that the length of the preceding boom and pre-recession values of the current account balance and credit volumes are correlated with post-recession revisions of potential output. We find however little evidence for the reverse causality hypothesis. Revisions of potential output estimates before recessions are neither significant nor substantial. Our results call for improved methods for estimating potential output and provide evidence against the existence of substantial hysteresis following demand shocks.

In Chapter 3, co-authored by Jonas Dovern, we analyze how component estimates of potential output are revised for 27 European countries after major economic crises. Expert revisions for these countries can be decomposed into revisions of the capital stock, trend labor, and trend total factor productivity (TFP). Differences in size and timing of component revisions indicate that potential output is revised through different channels. Revisions of capital stock estimates may signal a decline of investment in physical capital, revisions of trend labor estimates may indicate structural changes at the labor market, and revisions of trend TFP can be caused by a decline in R&D investment. The assessment of the relevance of each channel is important for policy makers to tailor well-defined policies.

Using real-time data from the European Commission, we analyze how estimates of the capital stock, trend labor, and trend TFP are revised downwards after the Great Recession in 2008 and the European sovereign debt crisis in 2011. We show that revisions to different components of potential output contribute equally to the substantial overall decline in estimates of potential output levels. Revisions of trend labor are thereby predominantly driven by revisions of the non-accelerating wage rate of unemployment. The heterogeneity of revisions across European countries after the Great Recession is large, suggesting that different policies are needed to bring countries back to their previous growth paths. For example, Central and Eastern European countries observe strong reductions of capital stock estimates as the main source of overall potential output revision whereas the sample of euro area countries, which were severely hit by the Great Recession, mainly face revisions of trend labor estimates.

In Chapter 4, I evaluate how the explicit modeling of different economic relationships can improve the reliability of output gap estimates for the United States. Specifically, I analyze whether credit growth helps to produce more reliable output gap estimates as the results from Chapter 2 and the findings by Borio et al. (2014, 2017) suggest. I outline a small unobserved components model which features several well-established macroeconomic relationships. First, I include Okun's law by linking the output gap to the cyclical component of unemployment. Second, a Phillips curve relationship connects the output gap to the cyclical component of inflation. Third, I assume that the cyclical component of the capacity utilization rate, a survey-based measure, contains information about the development of the output gap.

Estimating the outlined model and various alternative versions, I show that reliable estimates of the output gap do not depend on model complexity but the featured economic relationships. Models which include Okun's law produce more reliable estimates than models which solely feature a Phillips curve or a capacity utilization relationship. Output gap estimates from a model which additionally includes credit growth as an explanatory variable only improve marginally. In a forecast exercise, I however document that output gap estimates are significantly closer to ex-post estimates if the model is extended by credit growth.

Chapter 5, co-authored by David A. Vespermann, analyzes whether the effect of fiscal stimulus depends on the state of the economy. Specifically, we investigate the effectiveness of government spending in the Excessive Deficit Procedure (EDP). The EDP is the corrective arm of European fiscal governance and designed to ensure sustainable public budgets. The procedure is launched if a member country of the European Union runs an excessive deficit or its debt level exceeds a certain threshold. The European Commission and Council then continuously monitor the actions of the member state and suggest measures to reduce the fiscal imbalances. These measures are an additional strain on public balances and may limit the effectiveness of government spending.

Following Auerbach and Gorodnichenko (2012, 2013) and Ramey and Zubairy (2018), we estimate fiscal multipliers by state-dependent local projections for a panel of 17 European countries. We document that cumulative fiscal spending multipliers are larger for countries in the EDP. This result is driven by lower interest rates and substantial crowding-in of private investment in response to a positive government spending shock. Fiscal spending multipliers in the EDP are even larger in times of weak fiscal position or recessionary episodes, indicating that the procedure is particularly effective. We show that the EDP is not simply a proxy for these times and hence suitable to explain variations in fiscal spending multipliers. In addition, we find that policy makers underestimate fiscal multipliers in real time. The results suggest that the EDP is functional and increases the effectiveness of government spending.

In summary, this thesis focuses on the effects of fiscal policy and recessions on economic growth and potential output. It highlights that fiscal policy is especially effective in the EDP and that potential output estimates are substantially revised after recessions. Revisions of potential output estimates are mainly driven by supply shocks and previous measurement error. Extending unobserved components models by a variable like credit growth, which has explanatory power for the revisions of potential output estimates, helps to make output gap estimates more reliable.

2

Recessions and Potential Output: Disentangling Measurement Errors, Supply Shocks, and Hysteresis Effects[1]

Joint with Jonas Dovern

2.1 Introduction

Estimates of potential output (PO) are important for decisions about monetary and fiscal policy. Although PO estimates are meant to proxy the level of economic output that is sustainable in the long-run and independent of cyclical (demand-driven) fluctuations, they have been revised downwards in response to the Great Recession for many countries (see, e.g., Benati, 2012; Ball, 2014). This has renewed interest in the question of how sensitive PO estimates are to severe economic downturns and which factors might help to anticipate and avoid post-recession revisions of such estimates.

Against this background, this chapter addresses the following research questions: Do recessions have permanent effects on PO estimates? What are the reasons behind downward revisions of PO estimates in the aftermath of recessions? To empirically investigate these questions, we use a newly compiled data set with real-time vintages of the Economic Outlook (EO) provided by the Organisation for Economic Co-operation and Development (OECD) that allows us to trace the development of revisions to the *level* of PO after recessions.

Avoiding PO revisions (and policy mistakes caused by them) in the future requires a better understanding of the underlying reasons for PO revisions. We distinguish between four main explanations why PO estimates are revised downwards following a recession. If recessions are caused by permanent supply shocks, the revisions are a reflection of the lower-than-previously-expected long-run output path of the economy (explanation 1). If recessions are caused by demand shocks, revisions of PO estimates would indicate that the analysts believe

[1] This chapter is based on our article in *The Scandinavian Journal of Economics* (Dovern and Zuber, 2020a). It is available at https://doi.org/10.1111/sjoe.12385.

that hysteresis effects lead to permanent output effects (explanation 2). If PO is overestimated before a recession, subsequent revisions of PO estimates are "merely" a correction of previous measurement errors (explanation 3). Finally, causation could run in the opposite direction, as suggested in Blanchard et al. (2015) and Blanchard et al. (2017). In this case, declining income expectations induced by the PO revision lead to a fall in aggregate demand and a recession (explanation 4). Note that in this case, PO revisions should precede recessions.

This chapter presents three main findings regarding the revision of PO estimates. First, we document substantial post-recession revisions of the level of PO estimates which are predictable and partly result from pre-recession estimation errors. We also show that revisions happen very gradually over a period of up to five years. Second, we find little evidence that PO revisions precede recessions. Finally, supply shocks systematically lead to PO revisions while demand shocks prove to be unimportant for revisions of the level of PO.

We contribute to a number of strands in the literature. First, the chapter adds to the literature that investigates the reliability of PO estimates. Starting with Orphanides and van Norden (2002) and Orphanides (2003), the unreliability of such estimates in real time is documented in many studies (see, e.g., Orphanides et al., 2000; Camba-Mendez and Rodriguez-Palenzuela, 2003; Marcellino and Musso, 2011; Jacobs and van Norden, 2016), with Edge and Rudd (2016) presenting somewhat more optimistic findings for output gap estimates published by the Federal Reserve Board since the 1990s. A number of contributions suggest broadening the information set that is used to estimate PO to obtain more stable estimates. Garratt et al. (2008) show how one can use information from different data vintages and model data revisions explicitly to obtain more reliable PO estimates. More recently, Borio et al. (2014) and Borio et al. (2017) suggest using information about the financial cycle to improve estimates. The chapter expands on this literature as it is the first that systematically documents the instability of estimates of PO levels after recessions as well as the underlying reasons.

Second, the chapter relates to other studies that analyze whether recessions or financial crises affect *potential* output (estimates) or the corresponding growth rates. Based on OECD real-time data for 23 countries, Ball (2014) shows that PO estimates remain permanently below pre-recession trends after the Great Recession of 2008/09.[2] Haltmaier (2012) (using the Hodrick-Prescott (HP) filter) and Martin et al. (2015) (using exponential trends) use PO estimates obtained ex post by filtering the most recent data vintage. Both papers find that PO growth decreases permanently following recessions. Using the production function approach to estimate PO, Furceri and Mourougane (2012) document a similar effect for the times after financial crises for a sample of 30 OECD countries. Finally, Benati (2012) uses a structural

[2] Using PO estimates from real-time vintages of the IMF World Economic Outlook, Fatás and Summers (2018) provide evidence that fiscal consolidations contributed to the decline of PO during this period.

vectorautoregressive (VAR) model to provide evidence that PO growth slowed down after the Great Recession in the United States (US), the Euro area, and the United Kingdom. So far, none of these papers have used comprehensive real-time data on actual estimates of PO levels. Such real-time data are however necessary to understand the exact timing and causes of PO revisions.

The lack of use of real-time data is also a shortcoming of the third strand of literature that the chapter contributes to. This literature analyzes whether recessions or financial crises affect *actual* output or its growth rate. The most notable study in this context is by Cerra and Saxena (2008), who show that output losses following financial or currency crises are very persistent in the period between 1960 and 2001. Other studies, such as Papell and Prodan (2012) and Abiad et al. (2009), confirm these findings. Based on data for 100 financial crises over the last 150 years, Reinhart and Rogoff (2009, 2014) broaden the view and show that financial crises have negative impacts on a wide range of variables, such as asset prices, employment or government debt. Finally, a number of studies (see, e.g., Hosseinkouchack and Wolters, 2013; Blanchard et al., 2015) provide evidence that also regular recessions tend to permanently reduce the level of output.

In a wider context, this chapter is related to the literature on macroeconomic hysteresis effects, i.e., the notion that temporary shocks, such as monetary shocks or demand shocks, might have long-lasting or even permanent effects on output (Blanchard and Summers, 1986, 1987; Lindbeck and Snower, 1986; Stadler, 1986, 1990). We do not contribute to the discussion of different hysteresis mechanisms. But they are one potential reason (explanation 2) for downward revisions of PO estimates after recessions.

The remainder of this chapter is structured as follows. Section 2.2 explains our data and how we make PO estimates from different data vintages and for different countries comparable. Section 2.3 contains the empirical results of our study. It first presents non-parametric statistics that show when and by how much PO estimates are revised following a recession. It then provides evidence about potential causes behind the observed revisions; in particular, it contains evidence that supply and demand shocks trigger different patterns of PO revisions. Section 2.4 concludes.

2.2 Data

In this section, we describe how we identify recessions, present our real-time data set and explain how we extrapolate and normalize PO estimates.

2.2.1 Identification of Recessions

To identify recessions, we rely on the simple and transparent method proposed by Bry and Boschan (1971), as adapted for quarterly time series by Harding and Pagan (2002). The algorithm identifies business cycle peaks and troughs by searching local maxima and minima of the level of GDP while ensuring that local maxima and minima alternate and each phase and full cycle exceeds some minimum length.[3] We apply this algorithm to data on real gross domestic product (GDP) from the most recent vintage of the OECD Main Economic Indicator (MEI) database. In total, we identify 95 recessions between 1990 and 2017, for which we have corresponding data on PO estimates by the OECD.[4] The algorithm is accurate and yields plausible recession dates. For the US, for example, our recession dating coincides with the business cycle dates provided by the National Bureau of Economic Research (NBER) with respect to both start and end of the recessions.[5] The mean duration of the identified recessions is 4.5 quarters and the maximum loss in output (relative to the pre-recession peak) is −3.3% on average.

2.2.2 Real-Time Data from Economic Outlook

Our main data are different vintages of the EO published by the OECD in spring and autumn of each year. This source contains macroeconomic data for member states of the OECD along with forecasts (one and two years ahead) and estimates of unobservables (such as PO) made by the OECD. We use the OECD data for three main reasons. First, the OECD covers a large sample of countries for which it produces consistent PO estimates. Second, the OECD uses the production function approach to estimate PO (see Beffy et al., 2006) which is also widely used elsewhere. Finally, the availability of real-time data allows us to look at actual revisions of PO estimates and their timing.

Our sample of data vintages that contain information about PO estimates ranges from spring 1989 (EO No. 45) until spring 2018 (EO No. 103). It covers the 27 countries which are listed in Table 2.1. We have a full set of 59 vintages for 16 of those countries. For Greece, New Zealand, and Norway, only a small number of vintages is missing.

Our main variable of interest are the OECD's estimates of PO levels which we denote by $\bar{y}$.[6] In addition, we use information on the level of GDP (y), the current account balance (in % of

---

[3] We require each business cycle phase to last for at least two quarters and each complete cycle (trough to trough and peak to peak) for at least five quarters.

[4] We exclude one "recession" which is identified by the algorithm (Denmark 1990Q3–1991Q4) because it is clearly misclassified. For a full list of the recessions, see Table 2.6 in the Appendix.

[5] One exception is the NBER's call of a recession in 2001, which is not identified by our algorithm because it did not involve two consecutive quarters of negative GDP growth.

[6] Note that PO estimates and GDP for Norway refer to the level of domestic production excluding exploration of crude oil and natural gas, transport via pipelines and ocean transport.

Table 2.1—Sample overview of PO estimates

Country	1st vintage	# Vintages	Max. sample	# Recessions
Australia	1989–1	59	1961	1
Austria	1989–1	59	1961	4
Belgium	1989–1	59	1970	4
Canada	1989–1	59	1962	3
Chile	2012–1	13	1987	1
Czech Republic	2005–2	26	1992	2
Denmark	1989–1	59	1960	7
Finland	1989–1	59	1961	4
France	1989–1	59	1963	2
Germany	1994–1	49	1963	5
Greece	1989–2	58	1961	6
Hungary	2008–2	20	1992	2
Iceland	2000–1	37	1964	3
Ireland	1989–1	59	1961	2
Italy	1989–1	59	1960	7
Japan	1989–1	59	1962	7
Luxembourg	2005–2	26	1976	3
Netherlands	1989–1	59	1970	2
New Zealand	1989–1	54	1963	5
Norway†	1989–1	54	1965	4
Portugal	1994–2	48	1960	3
Slovenia	2011–2	14	1999	1
Spain	1989–1	59	1965	3
Sweden	1989–1	59	1964	3
Switzerland	1989–1	59	1961	7
United Kingdom	1989–1	59	1963	2
United States	1989–1	59	1960	2

Notes: "1st vintage" refers to the first vintage from which PO estimates are available. "Max. sample" notes the first year for which PO estimates are available in at least one vintage. We use data from the previous vintage to proxy missing vintages in the following cases: Greece (1991–1), Ireland (1991–1/1991–2), and Switzerland (1994–2). Five vintages of PO estimates are missing for New Zealand (1994–2 to 1996–2) and for Norway (1991–1 to 1993–1). PO estimates for unified Germany are not available in vintages before 1994–1. † Data on PO estimates for Norway refer to domestic production excluding exploration of crude oil and natural gas, transport via pipelines and ocean transport.

GDP), imports and exports (to construct a measure of trade openness), the level of public debt (in % of GDP), and the public primary balance (in % of GDP) from the EO vintages.

Because we use real-time data vintages, our data have a multi-dimensional structure. This allows us to track how the OECD's PO estimate for any particular year changes across vintages following the start of a recession. Consequently, we are not only able to analyze how large revisions are but also when they occur. We denote a variable x for country i and year t from vintage v by $x_{i,t}^{v}$.

A snapshot of the raw data is plotted in Figure 2.1 which shows all vintages of PO estimates for the US. The plot shows that revisions can be substantial. It is also evident from the figure that we need to normalize the data due to changes in national accounting standards and,

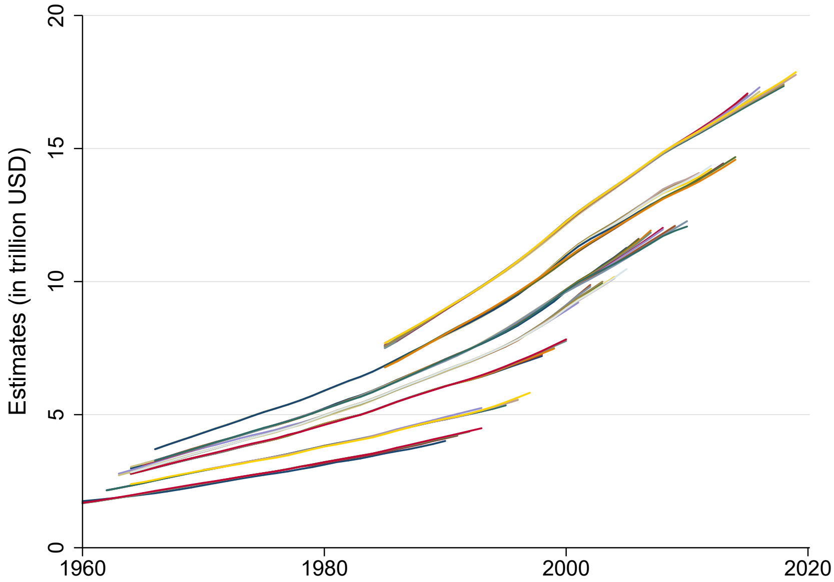

Figure 2.1. Raw data vintages of PO estimates for the US

Notes: The plot shows OECD estimates for (real) PO in the US from different EO vintages. Values are in trillions of real USD (with different base years).

most importantly, base years. This applies to all countries and we explain how we do this in Section 2.2.4.

2.2.3 Extrapolation of Potential Output Estimates

The OECD data contain PO estimates that reach two years ahead of the publication time, i.e., a vintage v' from a certain year t' contains information up to $\bar{y}^{v'}_{i,t'+2}$. Because we are also interested in medium-term revisions following a recession and would like to compare how the PO estimates for, say, the fifth year after a recession change during the recession and the following years, we need to extrapolate the raw OECD estimates.

We do so by expanding the OECD estimates for additional 10 years using the implied average potential growth rate of the last 10 observations (which include the OECD's forecast for the next two years). That is, we compute $\gamma = 1/10 \sum_{k=-7}^{2} \Delta \ln \bar{y}^{v'}_{i,t'+k}$, and obtain additional (log) PO estimates as $\ln \bar{y}^{v'}_{i,t'+k} = \ln \bar{y}^{v'}_{i,t'+2} + (k-2)\gamma$ for $k > 2$. Since these additional data points depend on our calculations and are no raw OECD estimates, we indicate below which results depend on the additional data and which results do not.

Given the high degree of smoothness of PO, our approach is adequate. The fit of a linear trend through the last 10 observations of the raw PO estimates is very good on average. Looking at the distribution of the corresponding R^2 across all vintages and countries in our sample reveals a median goodness of fit of above 0.99. In fact, even the 25$^{\text{th}}$ percentile of the distribution of R^2s is 0.99 and the 1$^{\text{st}}$ percentile is still 0.65. So there is little evidence that our extrapolations lead to large approximation errors (with respect to the unpublished long-run PO forecasts by the OECD). In addition, there is little reason to believe that those approximation errors, which presumably are mainly due to the fact that the OECD can anticipate demographic changes in real time, are systematically related to the occurrence of recessions. Thus, overall, we are confident that the extrapolation does not systematically affect our results although it might induce some noise.

2.2.4 Data Normalization

When comparing PO estimates from different vintages, we have to take into account potential changes in national accounting standards, base year, and/or unit of measurement (e.g., with the introduction of the euro). Since for a number of countries there are some vintages with samples that do not overlap, we cannot use a global base year to normalize all data. Instead, we use a different normalization for each identified recession, making comparable all vintages that are relevant for tracing revisions around this particular recession.

Denoting the first year of a recession by t_0 and the first vintage following the start of a recession by v_0, we construct normalized PO estimates using the following formula:

$$\tilde{y}^{v_0+k}_{i,t_0+s} = \bar{y}^{v_0+k}_{i,t_0+s} \times \frac{y^{v_0}_{i,t_0-s^*}}{y^{v_0+k}_{i,t_0-s^*}}, \tag{2.1}$$

where k ranges from $k_{min} < 0$ to $k_{max} > 0$ (determining the range of vintages around a particular recession that we consider), s ranges from $s_{min} < 0$ to $s_{max} > 0$ (determining the range of years around a particular recession that we consider), and $s^* > 0$ determines which year we use as the base year for the normalization. s^* needs to be sufficiently large to ensure that already the earliest vintage considered ($v_0 + k_{min}$) contains reliable data about GDP in year $t_0 - s^*$; otherwise, forecast errors made by the OECD or data revisions would influence our results. In practice, we set s^* equal to 5.

We apply a second normalization step to make sequences of PO estimates comparable across recessions and countries. More specifically, we normalize $\tilde{y}^{v_0+k}_{i,t_0+s}$ such that $\tilde{y}^{v_0}_{i,t_0} = 100$, i.e., the estimate of PO in the first year of a recession as reported in the first vintage following the start of the same recession is set to 100.

2.3 Empirical Results

In this section, we first analyze the timing and size of revisions to PO estimates. Second, we investigate factors which help to explain the size of these revisions and, third, we specifically distinguish between the revisions to PO estimates in demand- and supply-driven recessions.

2.3.1 Timing and Size of Revisions to Potential Output Estimates

We start by tracking how estimates for PO in a certain year evolve across successive vintages of the EO. Figure 2.2 shows the distribution of the evolution of PO estimates for the first and the fifth year following the start of a recession. The solid vertical line at 0 indicates the first data vintage after the start of a recession. Vintages to the right are published after the first recession vintage; vintages to the left before the recession. It is evident that, on average, PO is revised downwards in the aftermath of a recession.[7] This is a new finding since previous studies have focused on PO *growth*. The estimate for the first year after the start of a recession, for instance, is reduced by roughly 1.8%, on average, from v_0 to v_{10} (Table 2.2). Looking at later years shows that the gap between pre-recession PO estimates and the post-recession estimates increases with the distance to the recession start. This confirms evidence in Blanchard et al. (2015), who refer by "super hysteresis" to the phenomenon that PO does not only suffer from a one-off level shift but from a decrease in PO *growth* which causes the gap between PO levels and previous projections to increase over time. Revisions to PO estimates are made gradually over roughly five years following the recession; about 1/2 of the revisions are made until the end of the recession. The pattern of revisions seems to be non-monotonic in the sense that the initial downward revisions are partly reversed later on: the estimates from v_{20} are between roughly 0.1 percentage points (Year 5) and 2.3 percentage points (Year 10) higher than those from v_{10}.

Interestingly, PO estimates are not revised much, on average, before the start of a recession. This can be seen from the stable medians at the left side of both plots (for vintages up to v_0). Also for other years, revisions before recessions are very small compared to revisions after recessions (Table 2.2). Thus, the OECD does not anticipate significantly lower long-run PO levels already before recessions. Prima facie, this provides evidence against the hypothesis that the causation could run from downward revisions of PO to recessions (explanation 4) as implied by the theoretical model in Beaudry et al. (2017). In their model, revisions of overly optimistic views held about the growth path of an economy triggers recessions. The mechanics of the model are as follows: During a period in which agents overestimate the growth rate of total factor productivity too much capital is accumulated. Once the overoptimism is corrected, agents reduce their level of investment demand to reduce the capital stock. This leads to low demand, higher unemployment risk, and, in turn, to precautionary savings that lower demand even further. Important to note is that in this model—like in the arguments made in

---

[7] One problematic issue could be the secular decline in trend growth rates in most of the countries in our sample. If persistently unanticipated by analysts, this might lead to a pattern of downward revisions of PO across vintages irrespectively of recessions (or any other cyclical phenomena). To check that our results are indeed driven by the occurrence of recessions and not an "artefact" of this secular trend, we re-produced our results based on a set of "randomly distributed recessions" (see Appendix 2.A.2). These results do not exhibit the same patterns (the interquartile range easily covers the median level from v_0 for all years and vintages).

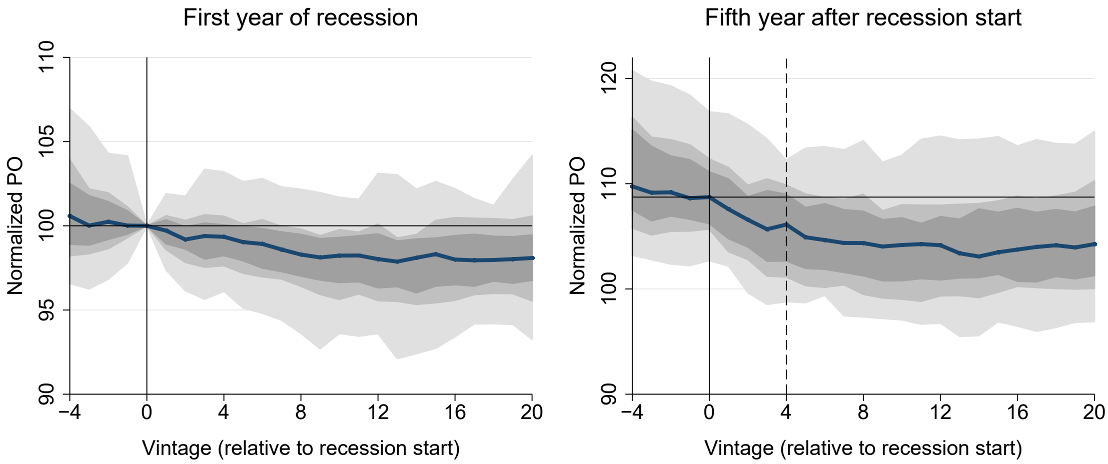

Figure 2.2. Revisions to PO estimates after recessions

Notes: The lines show the median revisions to OECD PO estimates for the first/fifth year after a recession start across different vintages. Values to the left of the dashed line depend on our extrapolation of the PO estimates. The data are normalized such that the value in the first recession year as estimated in the first vintage following the start of the recession is equal to 100. The sample includes 95 recessions. Grey shaded areas represent the 5th to 95th percentile range, the 17th to 83rd percentile range, and the interquartile range.

Blanchard et al. (2015) and Blanchard et al. (2017)—the recession starts only after views about PO are revised and, hence, we should observe downward PO revisions already before the start of a recession.[8]

Another prominent feature of our results is the large variation across countries and recessions. The band spanned by the 5th and the 95th percentile is large in all cases and ranges from solid positive numbers to substantially negative ones. After all, the huge variation is not too surprising given the very different situations of countries when hit by a recession.[9] Focusing on the 83rd percentile line shows that it plunges below the median level in vintage v_0 between roughly v_8 and v_{15} in both sub-plots. Thus, although the variation is large, we see a decline of PO estimates in the vast majority of cases.

Since economic structures and also economic policies have changed considerably over our sample period, it is a plausible hypothesis to suspect the results to be time varying. To explore how much of the heterogeneity can be explained by simple time fixed effects, we compare median PO revisions for different time periods. Because we do not observe many recessions in every year (sometimes none), we look at periods of five years. A benefit of this approach is that it forms "natural" clusters of recessions. For example, recessions associated with the Great Recession (2007–09) and recessions associated with the European debt crisis

[8] To the extent that consumers, or private sector agents in general, anticipate lower PO levels earlier than the OECD, the case that the demand short-fall is indeed driven by this anticipation remains, of course, possible.

[9] We come back to this issue in Section 2.3.2 and Section 2.3.3 where we analyze which factors can be used to explain differences in the size of PO revisions.

Table 2.2—Change of PO estimates around recessions

Year	$\tilde{y}_t^{v_0}$	Change vs. $\tilde{y}_t^{v_0}$ (in %) $\tilde{y}_t^{v_{-4}}$	$\tilde{y}_t^{v_{-2}}$	$\tilde{y}_t^{v_2}$	$\tilde{y}_t^{v_4}$	$\tilde{y}_t^{v_{10}}$	$\tilde{y}_t^{v_{20}}$
1	100.00	0.58	0.24	−0.82	−0.65	−1.77	−1.91
2	101.94	0.84†	0.40	−1.00	−1.11	−2.36	−2.25
3	104.09†	1.30†	0.31†	−1.73†	−1.84†	−2.95†	−2.73†
5	108.73†	0.94†	0.42†	−1.98†	−2.40†	−4.19†	−4.11†
10	121.17†	1.75†	0.10†	−2.62†	−3.91†	−7.13†	−4.80†

Notes: The full sample median is based on 95 recessions. $\tilde{y}_t^{v_j}$ refers to median of the normalized PO estimates in a particular year (indicated by the row label) after the start of a recession as reported in the vintage j relative to recession start. $j = 0$ refers to the first vintage following the start of a recession. Estimates for each recession are rescaled such that $\tilde{y}_1^{v_0} = 100$. † indicates that the computation of the result involves our extrapolation of PO estimates.

Table 2.3—Differences in recessions and PO revisions across time

	N	Length	Depth	PO revision to trough	v_{10}	p-value for Mood's median test 1995–99	2000–04	2005–09	2010–14
1990–94	19	4.1	−3.17	−1.08	−3.06	0.395	0.213	0.009***	0.516
1995–99	8	3.1	−0.95	−1.18	−2.32		0.284	0.011**	0.403
2000–04	13	3.8	−0.85	−1.63	−2.52			0.000***	0.619
2005–09	25	5.9	−6.87	−5.53	−8.89				0.000***
2010–14	25	4.4	−2.36	−1.94	−2.64				
2015–19	5	2.6	−1.06						
Total	95	4.5	−3.31	−1.94	−4.19				

Notes: "Length" states the average recession duration in quarters. "Depth" refers to the average deviation from the pre-recession peak level of output to the trough (in %). The column "PO revision to trough/v_{10}" displays the median PO revisions for the fifth year after a recession made between v_0 and the first vintage after a recession or v_{10} respectively. PO revisions for the time period 2015–19 are missing since they rely on vintages which are yet to be published. Mood's median test is used to test whether the subsample median revisions between v_0 and v_{10} are pairwise statistically different. ***, **, and * correspond to significance levels of 1%, 5%, and 10%, respectively.

(2010–12) fall into separate groups. It turns out that the characteristics of the recessions and the PO revisions differ greatly across the subsamples (Table 2.3). The average length of recessions varies, for instance, from 3.1 quarters (1995–99) to 5.9 quarter (2005–09). These two periods also bound the range of median PO revisions between the start of recessions and the 10th post-recession vintage. Overall, the subsample results give a first impression that longer and deeper recessions seem to be associated with larger downward revisions of PO.

We find that compared to other sub-periods revisions are, on average, significantly larger during the period covering the Great Recession. We obtain this result by applying Mood's median test to the PO revision between the recession start (v_0) and five years later (v_{10}). We prefer this test (over, e.g., the Wilcoxon-Mann-Whitney test) since it does not assume similar variances across subsamples. If the test rejects the null hypothesis of equal medians despite its low power, we are confident that medians are significantly different from each

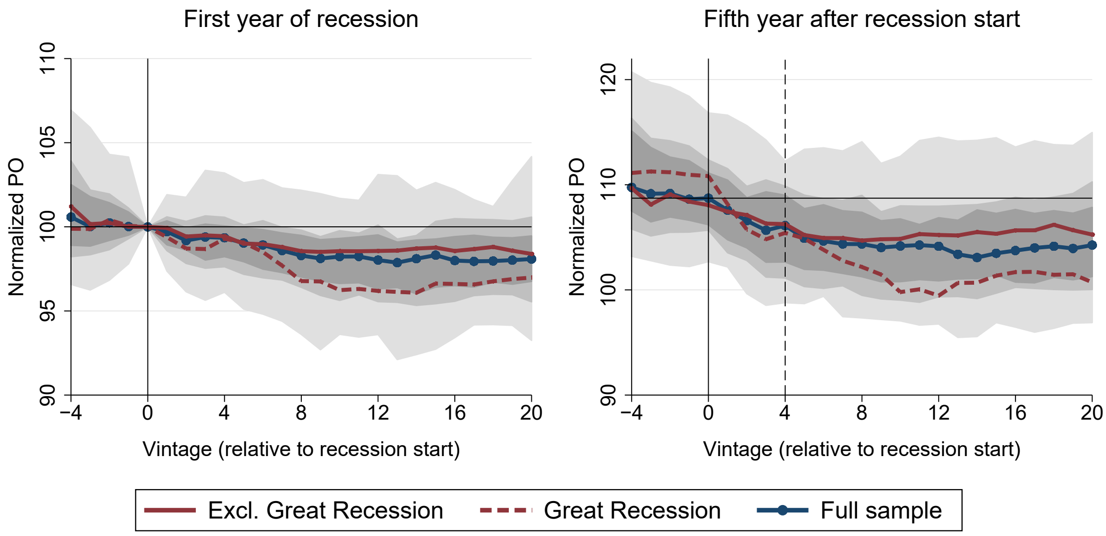

Figure 2.3. Revisions to PO estimates after recessions excluding the Great Recession

Notes: The lines show the median revisions to OECD PO estimates for the first/fifth year after a recession start across different vintages. We separate the sample into 71 recessions which begin before 2007 or after 2009 and 24 recessions which start in 2007–09. Values to the left of the dashed line depend on our extrapolation of the PO estimates. The data are normalized such that the value in the first recession year as estimated in the first vintage following the start of the recession is equal to 100. For comparison, we show the median based on the full sample. Grey shaded areas represent the 5th to 95th percentile range, the 17th to 83rd percentile range, and the interquartile range for the full sample.

other. Pairwise comparisons of periods not covering the Great Recession do not indicate any significant differences. This shows that the Great Recession has been exceptional in terms of its long-term damage to aggregate production capacities.

In light of this finding and to exclude the possibility that our full-sample results are heavily influenced by the Great Recession, we split our sample into one group of recessions that start during the years 2007–09 and one group consisting of all other recessions. Figure 2.3 shows that even though the median revision of PO estimates after the Great Recession is large relative to the historical norm, the median downward revisions for the remaining sample is almost as large as for the full sample, indicating that the identified phenomenon is not a special feature of the Great Recession and its aftermath.

Another way of presenting the data on PO revisions is to look at the long-run revision made to estimates for several years around the start of a recession. Figure 2.4 shows that the revisions (made over a span of ten years) to PO in years far after the start of a recession are much larger than for the first recession year. In addition, the data show that post-recession revisions are also made to PO in the years before recessions. PO for the fifth year before a recession, for instance, is revised downwards roughly 0.7% on average. Thus, because the OECD uses an estimation approach which is implicitly two-sided, it revises its PO estimates even for years up to a decade before a recession.

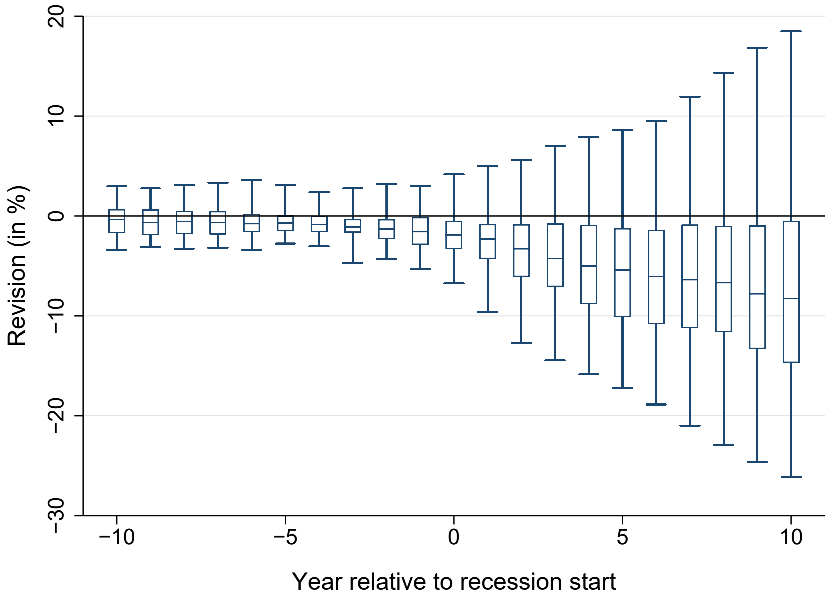

Figure 2.4. Revisions to PO estimates for different years around recession start

Notes: We compare PO revisions between the first and the 20$^{\text{th}}$ vintage after the recession start. In the Box-Whisker plot, the Whiskers include the 5$^{\text{th}}$ to 95$^{\text{th}}$ percentiles. The boxes correspond to the interquartile range and the medians are indicated by the horizontal lines inside the boxes.

As a benchmark for timing and size of the revisions, we compare how the revisions made by the OECD (using its preferred production function approach) relate to results one would have obtained in real time if one had applied a simple statistical filter to GDP data (incl. the OECD growth forecasts) to estimate PO for each data vintage. We use the one-sided Hodrick-Prescott (HP) filter (see, e.g., Stock and Watson, 1999) for such a purely statistical approach. We then extrapolate and normalize these alternative PO estimates in the same way as described in Sections 2.2.3 and 2.2.4.

It turns out that both size and timing of the OECD revisions are very similar to those which would have resulted from an application of the one-sided HP filter in real time—with the exception of estimates for the first year of a recession (Figure 2.5). The left plot indicates that five years after a recession started, the OECD, on average, revised downwards their estimate of PO in the first recession year by about 1.5 percentage points more than the application of the HP filter would suggest. The difference can be explained as follows: As seen above, the OECD implicitly applies a two-sided filtering method that causes the weak economic development during and after a recession to lower the trend estimate also for previous years. In contrast, the one-sided HP filter is only backward looking and estimates of PO in past years do only change due to data revisions. This becomes also evident if we compare the OECD revisions to those based on the two-sided HP filter which, on average, are more or less identical.

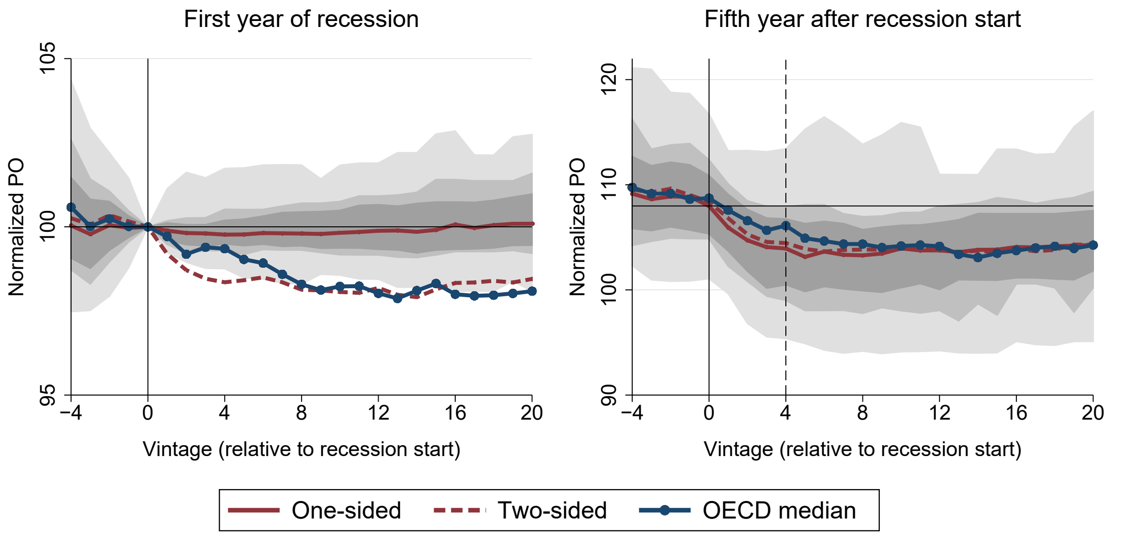

Figure 2.5. Comparison of OECD revisions to those based on the HP filter

Notes: The solid lines show the median revisions to PO estimates for the first/fifth year after a recession start across different vintages based on the one-sided HP filter. Values to the left of the dashed line depend on our extrapolation of the PO estimates. The data are normalized such that the value in the first recession year as estimated in the first vintage following the start of the recession is equal to 100. The sample includes 95 recessions. Grey shaded areas represent the 5th to 95th percentile range, the 17th to 83rd percentile range, and the interquartile range for the one-sided HP filter estimates. For comparison, we show the median based on the two-sided HP filter and the OECD median revisions (from Figure 2.2).

For the subsequent years, we do not observe remarkable differences between the size of the OECD's revisions and those based on the HP filters (results for the fifth year after a recession start are shown in the right plot of Figure 2.5).[10] The only observable difference is that the revision process at the OECD seems to be slightly more sluggish than that based on the filtering techniques, which can be explained by a smoothing tendency known from the macroeconomic forecasting literature (see, e.g., Nordhaus, 1987; Dovern and Weisser, 2011; Dovern et al., 2015). Overall, this suggests that a naive application of simple filtering techniques would speed up the revision process but would not change the size of revisions in the long-run.

2.3.2 Factors Explaining Potential Output Revisions

After establishing that PO estimates are, on average, substantially revised downwards following the occurrence of a recession, it is of interest whether we can identify variables—other than simple time fixed effects as in the previous section—that are correlated with the size of these revisions. This issue has not been systematically addressed in previous studies. However, in the long term, this might help to improve the quality of PO estimates.

[10] Results for other years look similar (in terms of the similarity of OECD and HP revisions) to those shown for the fifth year following the start of a recession.

We analyze the issue by regressing the size of the PO revision for a particular year (e.g., the first year of a recession) on a number of macroeconomic variables that might potentially correlate with our dependent variable. Thus, we reduce our data to a cross-sectional structure, in which every recession constitutes one observation.

More formally, we run regressions of the form:

$$\Delta\tilde{y}_{i,t_0+s}^{v'\to v''} = \beta X_{i,t_0} + \varepsilon_{i,t_0+s}\,, \tag{2.2}$$

where $\Delta\tilde{y}_{i,t}^{v'\to v''} = \ln\tilde{y}_{i,t}^{v''} - \ln\tilde{y}_{i,t}^{v'}$ denotes the revision of PO in country i for year t from vintage v' to vintage v'', X_{i,t_0} is a vector of covariates, β is a parameter vector of suitable dimension, and $\varepsilon_{i,t_0+s} \overset{iid}{\sim} N(0,\sigma^2)$ is an error term. Based on arguments drawn from previous empirical and theoretical contributions, we consider the following variables as elements of X_{i,t_0}: To characterize the nature of the recession, we include the length of the recession (in quarters) and the depth of the recession (in % of the peak level of GDP). Both are highly correlated and Blanchard (2018) argues that the degree of reallocation of input factors induced by recessions differs with the recession depth. To characterize the previous boom, we consider the length of the boom (in quarters) and the revision (made during the two years before the recession) of PO in the first year of the recession (in %). Beaudry et al. (2017) and Blanchard et al. (2017) argue that booms could be driven by overly optimistic views about the long-run potential of an economy, laying the foundation for later recessions and downward revisions of PO. Enders et al. (2018) empirically show that undue optimism contributes to output fluctuations. To measure a country's connectedness to the global economy, we include a pre-recession measure of trade openness (defined as the sum of exports and imports over GDP) and the pre-recession value of an index of financial openness (Chinn and Ito, 2006). Since the economy's ability to stabilize after macroeconomic shocks depends on the flexibility of the exchange rate and the policies of the central bank, we also consider the pre-recession values of measures of exchange rate stability and central bank independence (Aizenman et al., 2013). We also include the change of the current account balance (relative to GDP) over the two years before the recession (in percentage points). Darvas and Simon (2015) show how incorporating the current account balance in an unobserved components model yields more reliable PO estimates because the trade balance's ability to absorb excess demand is otherwise neglected. To measure the financial situation of the public sector and the private credit cycle, we include net public debt in the year before the recession in relation to GDP (in %), the public primary balance in the year before the recession in relation to GDP (in %), and the change in the ratio of credit to the non-financial sector to GDP over the two years before the recession (in percentage points).[11] Romer and Romer (2018) show that the output loss

[11] We obtain credit data from the Bank for International Settlements (BIS).

after financial crises is larger if a country lacks fiscal space to stabilize the economy. Borio (2014) and Borio et al. (2017) argue that credit booms lead periods of financial and economic distress. Finally, we include a number of indices measuring different aspects of economic flexibility and institutional quality (the Heritage Foundation Index of Economic Freedom, the World Bank's ease of doing business index, and two OECD measures of the strictness of employment protection laws). Blanchard et al. (2013) discuss the role of flexible (labor market) institutions for a country's ability to cope with macroeconomic shocks.

We present results for $s = 0$ and $s = 4$, i.e., we look at revisions of estimates for the first and the fifth year after a recession has started. In both cases, we look at revisions between the first vintage after the recession start and five years later ($v' = v_0$ and $v'' = v_0 + 10$).

The estimates in the left part of Table 2.4 indicate that the size of the post-recession revision to PO in the first recession year is significantly correlated with a number of factors. Given the low degrees of freedom, we re-estimate the equation by LASSO. The OLS and LASSO results are very similar in the sense that overall they identify the same variables to be relevant and also the size of estimates is often very similar. Therefore, we mainly focus on a description of the LASSO results for the model including the fixed effects. The depth of a recession seems to be a good predictor of subsequent revisions of PO, the elasticity being 0.06. We find also a linkage between pre-recession credit booms and current account changes to subsequent PO revisions. In the latter case, a 1 percentage point deterioration of the ratio of the current account to GDP goes along with an increase of PO revisions by 0.39 percentage points. This reflects that current approaches for estimating PO attribute substantial fractions of booms fueled by foreign credit to the structural ability of a country to sustain a certain output level—making later downward revisions of PO necessary.[12] A credit boom during the run-up to a recession is associated, on average, with a larger downward revision of PO. In line with results presented by Gadea Rivas and Perez-Quiros (2015), the linkage between credit booms and post-recession PO revisions disappears if we exclude data from the Great Recession (see Appendix 2.A.3), suggesting that this factor is only relevant if a recession is accompanied by the end of a pronounced financial cycle. Overall, based on the LASSO (OLS) estimates, the explanatory variables explain 41% (68%) of the variation of PO revisions for the first year of a recession.[13]

Turning to the post-recession PO revisions for the fifth year after the start of a recession shown in the right part of Table 2.4 reveals that we are able to explain an even larger share of the variation in this case. Interestingly, the set of strongly correlated variables is very similar

[12] The recent experience with Greece would be a typical example of this kind. Continuous current account deficits fueled a boom with a sectoral composition that could not be sustained once capital flows reversed after 2008.

[13] Note that the correlation with the measure of employment protection for temporary jobs is significantly different according to the OLS estimates but not selected as an important variable by the LASSO approach.

Table 2.4—Factors of PO revisions

Dependent variable:	$\tilde{y}_1^{v10} - \tilde{y}_1^{v0}$				$\tilde{y}_5^{v10} - \tilde{y}_5^{v0}$			
	OLS	OLS	LASSO	LASSO	OLS	OLS	LASSO	LASSO
Recession length	−0.019	−0.068			−0.142	−0.251		−0.075
	(−0.16)	(−0.55)			(−0.63)	(−1.08)		
Recession depth	0.132	0.023	0.153	0.059	0.415*	0.179	0.587	0.378
	(1.23)	(0.18)			(1.98)	(0.73)		
Length of previous boom	−0.012	−0.003	−0.005		−0.060***	−0.048*	−0.040	−0.015
	(−1.06)	(−0.20)			(−2.77)	(−2.00)		
$\tilde{y}_1^{v0} - \tilde{y}_1^{v-3}$	−0.187*	−0.153	−0.019		−0.268	−0.225		
	(−1.82)	(−1.37)			(−1.34)	(−1.07)		
Trade openness $(t-1)$	0.171	0.016			1.709	1.956		
	(0.24)	(0.02)			(1.25)	(1.28)		
Chinn-Ito index $(t-1)$	1.050*	1.163*	0.100		0.467	0.956		
	(1.74)	(1.76)			(0.40)	(0.77)		
Exchange rate stability $(t-1)$	0.157	0.388			−0.922	−0.595		
	(0.14)	(0.34)			(−0.41)	(−0.27)		
Monetary independence $(t-1)$	−0.107	0.640			−1.635	0.223		
	(−0.08)	(0.43)			(−0.63)	(0.08)		
Δ CA $(t-1)$	0.419***	0.394***	0.421	0.385	0.754***	0.657***	0.813	0.752
	(3.80)	(3.49)			(3.50)	(3.08)		
Primary balance $(t-1)$	0.042	0.085			0.069	0.078		
	(0.65)	(1.03)			(0.55)	(0.50)		
(Net) public debt $(t-1)$	0.013*	0.015**			0.034**	0.037**	0.001	
	(1.79)	(2.02)			(2.41)	(2.67)		
Δ Credit/GDP $(t-1)$	−0.039**	−0.038**	−0.025	−0.019	−0.104***	−0.112***	−0.073	−0.068
	(−2.37)	(−2.26)			(−3.29)	(−3.53)		
Economic Freedom $(t-1)$	0.081	0.114			0.064	0.107		
	(0.88)	(1.17)			(0.36)	(0.58)		
Ease of Doing Business index $(t-1)$	−0.010	−0.014			−0.083	−0.108*		
	(−0.35)	(−0.46)			(−1.47)	(−1.89)		
Employment protection (dismissals) $(t-1)$	0.013	0.019			−0.726	−0.808		−0.064
	(0.03)	(0.05)			(−0.91)	(−1.04)		
Employment protection (temporary) $(t-1)$	0.634*	0.795**			1.204*	1.708**		
	(1.79)	(2.15)			(1.74)	(2.44)		
Constant	−10.585	−13.361	−1.403	−1.118	−5.935	−10.999	−1.204	−1.121
	(−1.25)	(−1.54)			(−0.36)	(−0.67)		
5-years FE	No	Yes	No	Yes	No	Yes	No	Yes
N	70	70	70	70	70	70	70	70
R^2	0.65	0.68	0.36	0.41	0.75	0.78	0.52	0.56

Notes: Numbers in parenthesis are t-statistics. ***, **, and * correspond to significance levels of 1%, 5%, and 10%, respectively. 5-years FE group the recessions into the periods 1990–94, 1995–99, 2000–04, 2004–09, 2010–14. The LASSO tuning parameter is obtained by leave-one-out cross-validation. We indicate by $t-1$ that we use values of the variables from the year before the first recession year. The effective sample of recessions that we can use for the regressions is smaller than stated in Section 2.2.1 since relevant vintages are not available for recessions at the margins of our sample: We cannot use ten recessions that start in 2013 or later. Five additional recessions drop from the sample because the availability of vintages starts later (GRC–1990, HUN-2008, SVN–2011) or due to gaps in the available vintages (NZL–1991, NZL–1997). In two cases (CAN–1990 and FIN–1990), we use information from v_{-2} instead of v_{-3} to calculate the pre-recession revision. For ten additional recessions, data on net public debt is missing.

which is reassuring since it demonstrates the robustness of our results. The major differences are that now the recession length is substantially negatively correlated with the PO revisions and also the length of the previous boom is a significant predictor of post-recession revisions of PO which suggests that undue optimism spreads the longer a boom lasts (the longer the boom the stronger the following downward revisions).[14] In addition, we observe that the estimates are stronger in all cases. In particular, the elasticity with respect to the depth of a recession is roughly 1/3 (close to 0.6 if we do not control for time fixed effects), indicating that, on average, much of the fall in output during a recession is deemed as permanent by the OECD. This high persistence of output losses is in line with findings in Blanchard et al. (2015).

The fact that the length of the previous boom, the change in the current account balance, and the change in the credit volume—all of which are commonly neglected in the estimation of PO—are correlated with post-recession revisions of PO estimates and together explain around 1/2 of their variation suggests that pre-recession mismeasurement of PO is one part of the story (explanation 3).

The chapter has the modest aim of identifying (from an ex post perspective) what factors lead to downward revisions of PO after recessions. We do not intend in this chapter to provide a complete analysis of whether one can use these insights to improve PO estimates before and during recessions. Still it is instructive to check which of the correlations that we identify "survive" once we switch to the use of real-time data. Results for a replication of Table 2.4 based on real-time data for trade openness, the change in the current account balance, the primary balance, the public debt ratio, and the change of the credit-to-GDP ratio are presented in the Appendix (Table 2.8). Note that we lose many observations because for some variables real-time data are not available for our entire sample—which is why we replicate also the results based on final vintage data for the reduced sample. With some exceptions, the results are very similar. In particular, we still find that information on the current account balance and credit volume is helpful. The correlation with the latter is not significant anymore in the OLS regressions but the LASSO still chooses it in the case of the revisions to PO in the fifth year after the recession.

2.3.3 Demand- vs. Supply-Driven Recessions

So far, we have not distinguished between different types of recessions. In general, however, we expect the need to revise downward the long-run potential capacity of an economy to be

[14] We also find a negative correlation between the degree of employment protection for individual dismissals in regular contracts. However, the sign of the effect is the opposite of that corresponding to the employment protection for fixed term contracts (that is significantly different from zero according to the OLS estimates) which leads us to conclude that the nexus between employment protection and post-recession revisions of PO estimates is not clear.

larger after a (permanent) negative supply shock than after a (temporary) negative demand shock. To see whether we can confirm this conjecture, we use two complementary approaches: First, we split the sample into supply-driven recessions and demand-driven recessions and compare the median revisions in each subsample non-parametrically. Second, we construct macroeconomic supply and demand shocks and estimate the response of PO revisions to each of these shocks.

To split our sample into a group of recessions that are more likely to be driven by supply shocks and one group of recessions that are more likely to be driven by demand shocks, we follow Blanchard et al. (2015). Of course, any classification is imperfect. Still, we present results for two different approaches (both of which we take from Blanchard et al. (2015)) because we think that they provide interesting tentative insights. First, we label all recessions as supply-driven that are associated with financial crises as identified by Laeven and Valencia (2013).[15] Second, we label all recessions as supply-driven that are associated with an increase of inflation in the first year of the recession (relative to the previous year). The idea here is that only supply-side shocks can lead to a decrease in production while at the same time increasing the inflation rate.

To get an intuitive idea about the implications of these two kinds of recessions, we start by presenting two examples. Consider the supply-driven recession in Japan in 1997 and the demand-driven recession in Germany in 2001.[16] We plot the development of labor productivity, real GDP per capita and PO estimates around those two recessions in Appendix 2.A.4. It is evident that productivity as well as GDP per capita remain well below their pre-recession trends in the case of the supply-driven recession in Japan while both return to their pre-recession trends in the case of the demand-driven recession in Germany. Also, the difference between the revisions to PO estimates is large: PO revisions for Japan after 1997 exceed the analogous revisions for Germany after 2001 by a factor of almost three.

Our results suggest that such differences between supply- and demand-driven recessions are a general stylized fact that holds for the entire sample. We focus on the revisions of PO estimates for the fifth year after the recession start. The results based on both classification schemes are similar, even though the differences between the median revisions for the two recession groups differ depending on which classification we look at (Figure 2.6). We observe

---

[15] Some might argue that financial crises are a demand phenomenon and their occurrence should not be used to identify supply-driven recessions. We agree, however, with Blanchard et al. (2015) that the factors causing the recession in those cases (e.g., reduction in credit supply, increasing risk premia, less efficient allocation of capital across sectors) predominantly and persistently affect aggregate supply.

[16] Before 1997, Japanese commercial banks had engaged in lending to other Asian countries on a broad scale, not least because of the low interest rates in Japan. Following currency shocks to Japan's surrounding countries, many foreign borrowers defaulted on their liabilities which lead to huge capital losses of Japanese banks as documented by Corsetti et al. (1999). Japan slipped into an "imported" financial crisis (OECD, 1998; Laeven and Valencia, 2013) that impaired the functioning of domestic credit markets and led to a recession. In contrast, the German recession in 2001 was due to "weakness of domestic and external demand" (OECD, 2001) driven by low export growth and a short-fall in domestic investment.

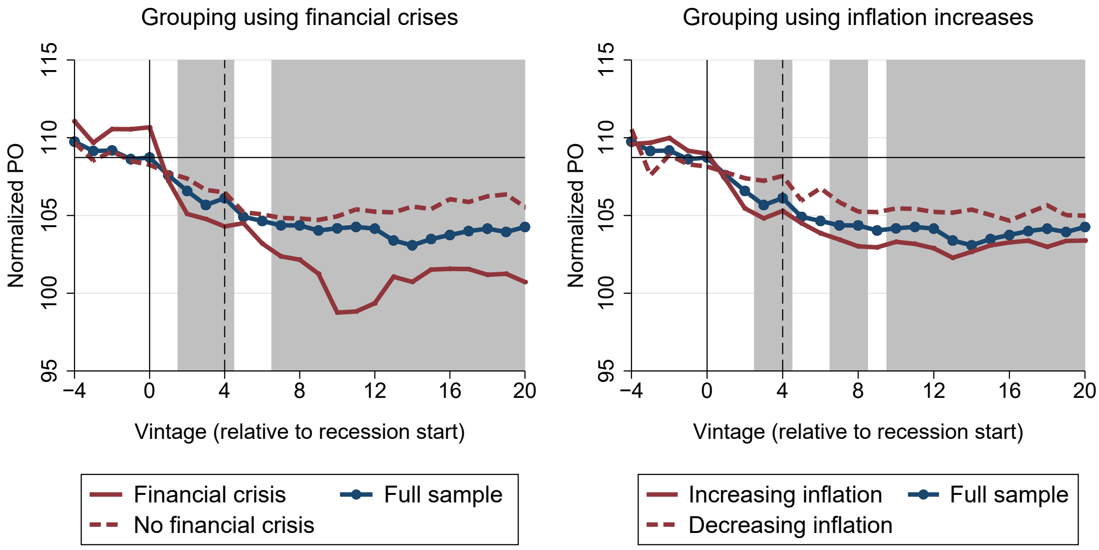

Figure 2.6. Effects of supply-driven and demand-driven recessions

Notes: The lines show the median revisions to OECD PO estimates for the fifth year after a recession start across different vintages. Financial crises correspond to the 21 crises identified by Laeven and Valencia (2013) which coincide with our sample. Increasing inflation means that annual inflation in the first recession year is higher than in the previous year (for 53 recessions). The two subsample medians are significantly different for the grey shaded vintages at the 10% level according to Mood's median test. Values to the left of the dashed line depend on our extrapolation of the PO estimates. The data are normalized such that the value in the first recession year as estimated in the first vintage following the start of the recession is equal to 100. The full sample includes 95 recessions.

the following set of consistent results. First, we find strong and persistent differences in the average size of revisions when contrasting recessions that are more likely to be driven by supply-side shocks to those driven mainly by demand shocks. Recessions driven by supply shocks, on average, are followed by (much) larger revisions to PO relative to demand-driven recessions. The subsample medians are significantly different from each other in both cases for most of the vintages after the recession start. Second, also after recessions that are likely to be driven by demand shocks the PO revisions are not equal to zero but substantially negative, suggesting that also these recessions leave permanent scars on estimates of the long-run growth paths. At this point, we cannot definitely say whether this is due to hysteresis effects and/or supply shocks that occur also during those recessions. Finally, the fact that the two subsample medians are very similar for vintages from before the start of the recessions is reassuring in the sense that it indicates that the differences are truly due to the different types of recessions.

The fact that PO is also revised downwards following recessions that we label as predominantly demand-driven does not necessarily imply that PO is revised in response to demand shocks; it could as well be a response to supply shocks that occur simultaneously. To investigate how PO estimates react to both types of structural shocks, we identify three types of such shocks and make inference about the impulse responses of PO estimates through local projections (Jordà, 2005) of cumulative revisions of PO estimates onto those structural shocks.

We consider one supply and two demand shocks. We identify labor productivity shocks as in Coibion et al. (2018). The shocks are computed as the residuals of autoregressive models (of order 4) for the change in labor productivity (that we obtain from the OECD), i.e., measuring unpredictable changes in labor productivity. We identify fiscal spending shocks as suggested by Auerbach and Gorodnichenko (2012). Using information from the EO data vintages, we measure the shocks by the difference between observed government spending (consumption plus—when available—investment) and the level that was expected half a year before. This is a very intuitive measure of unexpected changes in fiscal spending. Finally, we identify monetary policy shocks based on recursively identified structural VAR models. As in Coibion et al. (2018), we include real GDP growth, inflation, the unemployment rate and the short-term interest rate at a quarterly frequency (that we obtain from the OECD), using data from 1980Q1 until 2017Q4 or as available and assuming a VAR(4) specification. While the first shock constitutes a supply shock, the two latter ones are demand shocks. We aggregate all shocks to a semi-annual frequency to match the frequency of our vintage data. We denote the semi-annual observations of the three shocks as $\varepsilon_{i,t}^{lp}$, $\varepsilon_{i,t}^{fp}$, and $\varepsilon_{i,t}^{mp}$, respectively. Note that we "flip" the monetary policy shock so that positive values indicate an expansionary shock.

We compute cumulative revisions of PO estimates as $\Delta^h \tilde{y}_{i,y(t)}^t = \ln \tilde{y}_{i,y(t+h)}^{t+h} - \ln \tilde{y}_{i,y(t+h)}^t$, where $\tilde{y}_{i,y(t+h)}^t$ denotes, for instance, the PO estimate for country i from the vintage corresponding to the half-year t for the year that includes the half-year $t + h$.[17] $\Delta^h \tilde{y}_{i,y(t)}^t$ measures how much the estimate of PO for the year that lies (approximately) $h/2$ years ahead is revised between period t and h vintages later. We look at cumulative revisions because, as shown in Section 2.3.1, the revision process seems to be a gradual one. The local projections are given by:

$$\Delta^h \tilde{y}_{i,y(t)}^t = \alpha_i^h + \gamma_t^h + \beta^h \varepsilon_{i,t}^{\bullet} + e_{i,t}^h \,, \tag{2.3}$$

where we include country and time fixed effects. The sequence of β^hs provides the impulse response functions (IRF) of cumulative PO revisions to the different shocks.[18] For comparison, we run the same regressions for revisions of GDP forecasts.[19]

Figure 2.7 shows the results for horizons up to $h = 6$. As expected, PO estimates seem to respond positively to productivity shocks. In contrast, we do not find any significant responses to the two demand shocks; neither fiscal nor monetary policy shocks seem to systematically

---

[17] Note that the "target year" changes only every second time as we increase the horizon in steps that are determined by the semi-annual frequency of our vintage data.

[18] We also run the same regressions including additional lags of the structural shocks. The corresponding IRFs are almost unchanged and the conclusions do not change at all.

[19] Because we do not extrapolate these forecasts, the maximum h that we can consider in this case is rather low.

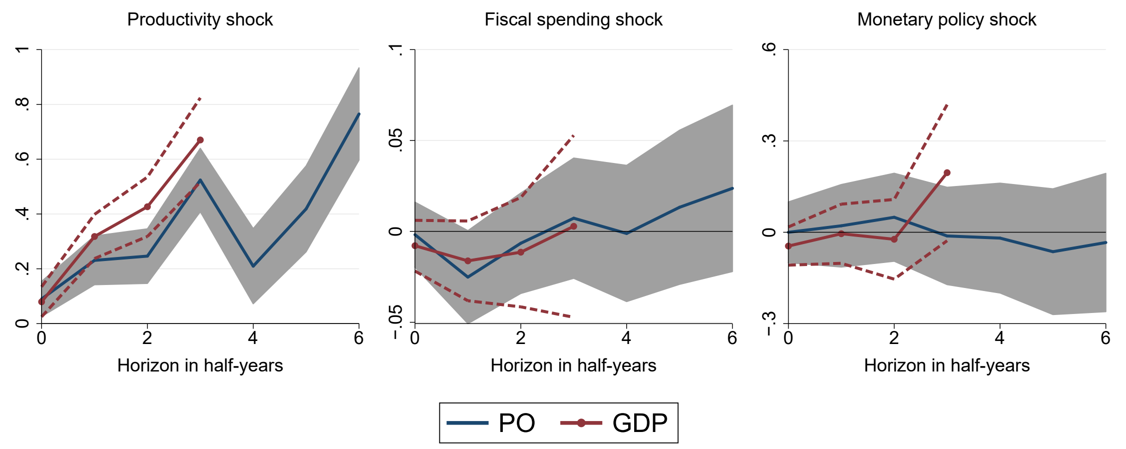

Figure 2.7. Response of PO estimates and GDP forecasts to selected structural shocks

Notes: The lines show the IRFs of (cumulative) revisions to OECD PO estimates and GDP forecasts to selected structural shocks. 90% confidence intervals are computed based on +/− 1.645 standard errors.

move PO estimates during subsequent years. Hence, in terms of PO *level* estimates we cannot confirm the "over-cyclicality" in response to demand shocks that Coibion et al. (2018) report for PO *growth* estimates. We find almost unchanged results if we exclude data from the years 2007–09 to check if our results are driven by the Great Recession. Hence, overall, our results support the conventional view that demand shocks have no long-lasting effect on the level of production.

Comparing the IRFs for PO and GDP forecasts, we see that GDP forecasts tend to react stronger to productivity shocks relative to PO estimates. For the two demand shocks we do not find any differences. Here, also GDP forecasts do not significantly react to the shocks. The insignificant responses of GDP forecasts to both fiscal policy shocks and monetary policy shocks suggest that either the OECD does not believe that they have any effect on GDP in the short-term or the OECD is not able to observe those shocks in real time.[20] Given how difficult it is to estimate policy impulses in real time, we think that the second explanation is probably the most likely one.

2.4 Conclusion

In this chapter, we have analyzed how OECD estimates for the *level* of PO are revised in the aftermath of recessions. We view these estimates as representative for a wide range of PO estimates published and used by policy institutions. Our empirical results are informative with respect to the plausibility of the four potential explanations for post-recession revisions

[20] A third explanation is, of course, that our approaches for estimating these shocks yield estimates that are subject to high measurement errors that make it difficult to identify the IRFs.

of PO levels that are discussed above (Table 2.5). We document that they tend to be revised downwards substantially and are broadly in line with results obtained by simple statistical filters. The revisions occur gradually over a period of approximately five years following the start of a recession. In addition, we find that revisions after supply-driven recessions are larger, on average, than those after recessions that are likely to be mainly triggered by adverse demand shocks. Since we do not find a significant response of PO estimates to structural demand shocks, we conclude that the PO revisions following demand side recessions are most likely due to simultaneous adverse supply shocks. Overall, the results suggest that permanent supply shocks (explanation 1) are more relevant for the PO revisions than demand shocks that lead to hysteresis (explanation 2).[21]

We identify a number of variables whose pre-recession values are correlated with the size of post-recession PO revisions. The correlation of revisions with credit booms, the current account development, and the length of the preceding boom suggest that a substantial part of the observed post-recession revisions is due to a previous underestimation of the boom/overestimation of PO (explanation 3). In our sample, we find no evidence that revisions to PO Granger cause recessions (explanation 4), rejecting the hypothesis of Blanchard et al. (2017) that "the anticipation of a less bright future is leading to temporarily weaker demand" (p. 639).

Our results have important policy implications. On the one hand, in the light of post-recession revisions to PO estimates, monetary and fiscal policy have to consider that the need for stimulative action after economic crises is smaller than indicated by pre-recession estimates. On the other hand, if very deep economic crises lead to substantial permanent output losses (as suggested by the large explanatory power of the recession depth for subsequent PO revisions), there would be a strong case for more aggressive stabilization policy during economic crises to mitigate the detrimental long-run effects (see also Erceg and Levin, 2014; Blanchard et al., 2015; Galí, 2016). In fact, Romer and Romer (2018) show that stabilization policy can help to mitigate the long-term effects of financial crises.

Our findings suggest a number of directions for future research. First, it would be interesting to look at the size and timing of post-recession revisions to estimates of the components of PO, i.e., potential labor input, the capital stock, and the trend of total factor productivity.[22] This would be informative about the mechanisms through which macroeconomic shocks lead to permanent output effects and could help developing future macroeconomic models. Second, the effect of macroeconomic stabilization policy during recessions on their long-term PO effects—similar in spirit to the analysis for financial crises in Romer and Romer (2018)—

[21] Please note the possibility that other types of demand shocks, which we do not consider here and which might lead to hysteresis effects, could be relevant during some recessions.

[22] Note that this is (not yet) possible based on the EO data because the number of vintages that contain such information is relatively small since the OECD started to include these variables not before the mid 2000s.

Table 2.5—Evidence on the potential explanations for post-recession revisions of PO

Explanation	Arguments for and against each explanation
1 Supply-side effects	+ Larger PO revisions after financial crises and when inflation increases in the first year of a recession + Significant PO reaction in response to productivity shocks
2 Hysteresis effects	– No PO revisions in response to fiscal/monetary policy shocks + Permanent PO revisions also after recessions that are likely to be demand-driven
3 Previous mismeasurement	+ Pre-recession macroeconomic conditions explain significant share of PO revisions
4 Reverse causality	– No PO revision before recession

Notes: Results marked by "+" support a particular explanation; results marked by "–" are evidence against a particular explanation.

requires a thorough investigation that is beyond the scope of this chapter. Finally, there seems to be room for improving the approaches that are used to estimate PO; taking a broader set of macroeconomic indicators into account could potentially lead to more stable estimates that are less prone to revisions. A first attempt into this direction has been made by a number of papers that try to identify unsustainable growth episodes by taking financial data into account when estimating PO (see, e.g., Borio et al., 2014, 2017). Our results support this idea. In particular, they suggest that taking data on international capital/trade flows into account might help to improve PO estimates.

2.A Appendix

2.A.1 List of Identified Recessions

Table 2.6—Identified recessions

#	Country	Start	Length	Depth	#	Country	Start	Length	Depth
1	AUS	1991q1	2	−1.47	49	GRC	2004q4	2	−0.94
2	AUT	1992q4	2	−0.54	50	GRC	2007q3	26	−27.44
3	AUT	2001q1	2	−0.38	51	GRC	2014q4	4	−2.38
4	AUT	2008q2	5	−5.09	52	HUN	2008q3	7	−7.70
5	AUT	2012q2	4	−0.98	53	HUN	2012q1	2	−2.48
6	BEL	1992q2	4	−2.88	54	IRL	2008q1	8	−10.68
7	BEL	2001q1	4	−0.38	55	IRL	2012q3	3	−1.73
8	BEL	2008q3	4	−3.81	56	ISL	2008q1	9	−12.95
9	BEL	2012q2	4	−0.67	57	ISL	2012q1	2	−2.37
10	CAN	1990q2	4	−3.43	58	ISL	2014q4	2	−0.72
11	CAN	2008q4	3	−4.48	59	ITA	1992q2	6	−1.50
12	CAN	2015q1	2	−0.34	60	ITA	1998q1	4	−0.60
13	CHE	1990q3	4	−1.22	61	ITA	2001q2	4	−0.73
14	CHE	1992q2	3	−1.70	62	ITA	2003q1	2	−0.61
15	CHE	1995q1	2	−0.45	63	ITA	2008q2	5	−7.95
16	CHE	1996q2	2	−0.50	64	ITA	2011q3	7	−5.20
17	CHE	1998q4	2	−0.27	65	ITA	2013q4	3	−0.16
18	CHE	2002q2	4	−0.87	66	JPN	1993q2	2	−1.35
19	CHE	2008q4	3	−3.36	67	JPN	1997q2	8	−2.31
20	CHL	2016q2	4	−0.48	68	JPN	2001q2	3	−1.83
21	CZE	2008q4	3	−5.80	69	JPN	2008q2	4	−8.69
22	CZE	2012q1	5	−1.92	70	JPN	2010q4	3	−2.75
23	DEU	1995q4	2	−1.05	71	JPN	2012q2	2	−1.04
24	DEU	2001q3	3	−0.48	72	JPN	2014q2	2	−1.85
25	DEU	2002q4	10	−0.51	73	LUX	2008q2	5	−8.10
26	DEU	2008q2	4	−6.93	74	LUX	2011q2	4	−1.72
27	DEU	2012q4	2	−0.67	75	LUX	2015q2	2	−2.39
28	DNK	1992q4	3	−2.04	76	NLD	2008q3	4	−4.52
29	DNK	1997q3	2	−0.30	77	NLD	2011q2	7	−2.04
30	DNK	2001q4	3	−0.24	78	NOR	2002q3	4	−0.81
31	DNK	2006q3	4	−1.05	79	NOR	2008q1	6	−2.70
32	DNK	2008q1	6	−7.07	80	NOR	2010q2	2	−3.52
33	DNK	2011q3	6	−0.50	81	NOR	2016q2	2	−1.02
34	DNK	2017q2	3	−1.08†	82	NZL	1991q1	2	−4.27
35	ESP	1992q2	5	−2.81	83	NZL	1997q4	3	−2.11
36	ESP	2008q3	6	−4.62	84	NZL	2000q2	4	−0.89
37	ESP	2010q4	12	−5.72	85	NZL	2008q1	5	−2.64
38	FIN	1990q2	13	−11.92	86	NZL	2010q3	2	−2.41
39	FIN	2008q1	6	−9.97	87	PRT	2002q2	5	−2.41
40	FIN	2012q2	4	−2.65	88	PRT	2008q2	4	−4.33
41	FIN	2013q4	6	−1.61	89	PRT	2010q4	9	−8.06
42	FRA	1992q2	4	−1.13	90	SVN	2011q3	7	−4.65
43	FRA	2008q2	5	−4.00	91	SWE	1991q1	9	−5.55
44	GBR	1990q3	5	−1.99	92	SWE	2008q1	5	−7.43
45	GBR	2008q2	5	−6.13	93	SWE	2011q4	5	−1.16
46	GRC	1990q2	2	−9.44	94	USA	1990q4	2	−1.32
47	GRC	1992q2	4	−4.69	95	USA	2008q1	6	−4.24
48	GRC	1994q4	2	−0.90					

Notes: "Length" states the duration of a recession in quarters. "Depth" refers to the deviation from the pre-recession peak level of output to the trough (in %). † indicates that a recession is ongoing at our sample end.

2.A.2 Simulation Based on Randomly Distributed "Recessions"

We randomly select 95 periods across countries and time for which we redo our analysis as if recessions had occurred in these periods. By repeating this, we create 10,000 data sets of randomly timed recessions. First, we calculate the median revision and percentile ranges for each data set. Then, we take the average of all 10,000 median revisions and percentile ranges to produce the figures below. On average, 3.7% of the random recessions start at the same time as the identified recessions.

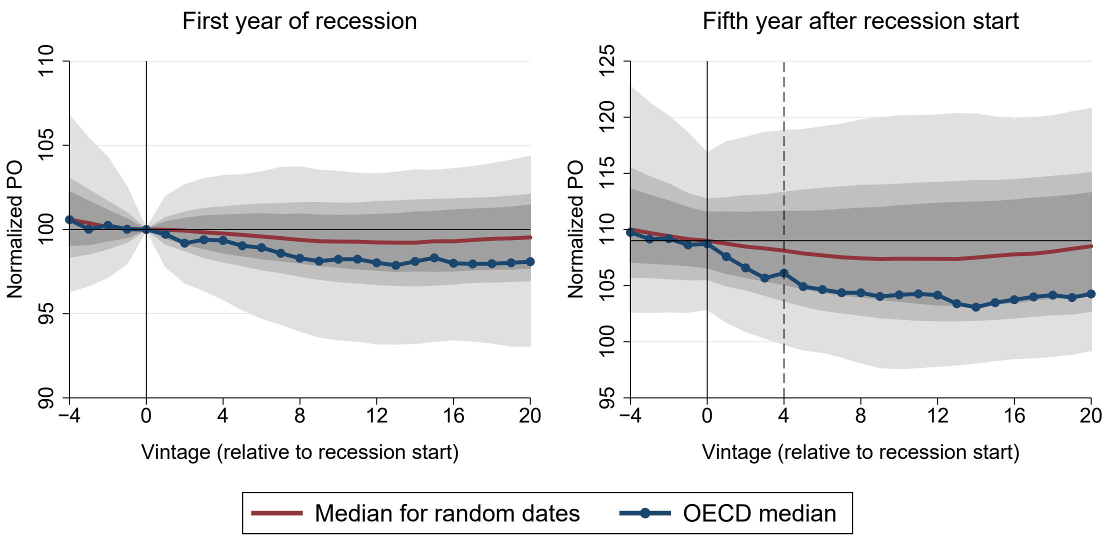

Figure 2.8. Revisions to PO during random periods

Notes: The lines show the revisions to OECD PO estimates for the first/fifth year after randomly selected periods across different vintages. Values to the left of the dashed line depend on our extrapolation of the PO estimates. Grey shaded areas represent the 5^{th} to 95^{th} percentile range, the 17^{th} to 83^{rd} percentile range, and the interquartile range, calculated as means across 10,000 random simulations. For comparison, we show the OECD median based on actual recessions (from Figure 2.2).

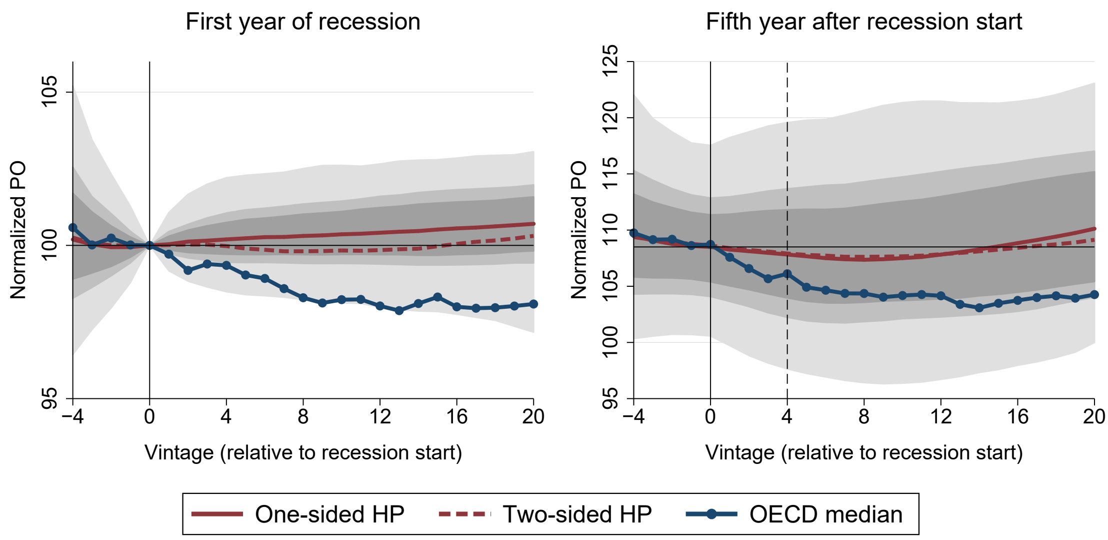

Figure 2.9. Revisions based on the HP filter during random periods

Notes: The lines show the median revisions to PO estimates for the first/fifth year after randomly selected periods across different vintages based on the one-sided HP filter. Values to the left of the dashed line depend on our extrapolation of the PO estimates. Grey shaded areas represent the 5^{th} to 95^{th} percentile range, the 17^{th} to 83^{rd} percentile range, and the interquartile range for the one-sided HP filter estimates, calculated as means across 10,000 random simulations. For comparison, we show the median based on the two-sided HP filter for randomly selected periods and the OECD median based on actual recessions (from Figure 2.2).

2.A.3 Robustness of Regression Results

Table 2.7—Factors of PO revisions excluding the Great Recession

Dependent variable:	$\tilde{y}_1^{v_{10}} - \tilde{y}_1^{v_0}$		$\tilde{y}_5^{v_{10}} - \tilde{y}_5^{v_0}$	
	OLS	LASSO	OLS	LASSO
Recession length	−0.219	−0.081	−0.480*	−0.326
	(−1.33)		(−1.71)	
Recession depth	−0.150		−0.291	0.023
	(−0.71)		(−0.80)	
Length of previous boom	0.036		0.011	
	(1.07)		(0.19)	
$\bar{y}_1^{v_0} - \bar{y}_1^{v_{-3}}$	−0.204		−0.340	
	(−1.44)		(−1.40)	
Trade openness $(t-1)$	0.586		4.174**	
	(0.59)		(2.45)	
Chinn-Ito index $(t-1)$	1.446*		1.508	
	(1.88)		(1.14)	
Exchange rate stability $(t-1)$	0.957		−0.911	
	(0.60)		(−0.33)	
Monetary independence $(t-1)$	1.999		1.947	
	(0.87)		(0.49)	
Δ CA $(t-1)$	0.412**	0.455	0.558**	0.862
	(2.75)		(2.18)	
Primary balance $(t-1)$	−0.018		−0.170	
	(−0.18)		(−0.98)	
(Net) public debt $(t-1)$	0.014		0.041**	0.004
	(1.61)		(2.75)	
Δ Credit/GDP $(t-1)$	−0.021		−0.055	−0.000
	(−0.47)		(−0.72)	
Economic Freedom $(t-1)$	0.046		−0.102	
	(0.36)		(−0.46)	
Ease of Doing Business index $(t-1)$	−0.040		−0.198***	
	(−0.97)		(−2.82)	
Employment protection (dismissals) $(t-1)$	−0.188		−1.661	−0.267
	(−0.33)		(−1.69)	
Employment protection (temporary) $(t-1)$	0.850*		1.892**	
	(1.93)		(2.51)	
Constant	−9.795	−1.044	2.391	−1.069
	(−0.86)		(0.12)	
5-years FE	Yes	Yes	Yes	Yes
N	48	48	48	48
R^2	0.67	0.33	0.73	0.42

Notes: We report the regressions from Table 2.4 for a sample without the Great Recession. Specifically, we exclude 24 recessions which start in 2007–09. The LASSO tuning parameter is obtained by leave-one-out cross-validation. Numbers in parenthesis are t-statistics. ***, **, and * correspond to significance levels of 1%, 5%, and 10%, respectively.

Table 2.8—Factors of PO revisions based on current-vintage data compared to final-vintage data

Dependent variable:	$\tilde{y}_1^{v10} - \tilde{y}_1^{v0}$				$\tilde{y}_5^{v10} - \tilde{y}_5^{v0}$			
	Current Vintage		Final Vintage		Current Vintage		Final Vintage	
	OLS	LASSO	OLS	LASSO	OLS	LASSO	OLS	LASSO
Recession length	−0.075	−0.053	−0.167	−0.089	−0.178	−0.203	−0.416	−0.281
	(−0.51)		(−1.22)		(−0.57)		(−1.55)	
Recession depth	0.284*	0.179	0.103	0.106	0.690*	0.513	0.423	0.364
	(1.70)		(0.64)		(1.97)		(1.33)	
Length of previous boom	0.005		0.008		−0.027		−0.024	
	(0.38)		(0.59)		(−1.00)		(−0.95)	
$\tilde{y}_1^{v0} - \tilde{y}_1^{v-3}$	−0.247	−0.067	−0.342**	−0.112	−0.216		−0.384	−0.002
	(−1.58)		(−2.50)		(−0.66)		(−1.43)	
Trade openness† $(t-1)$	−1.366		−0.840		−0.347		0.347	
	(−1.49)		(−0.94)		(−0.18)		(0.20)	
Chinn-Ito index $(t-1)$	−0.313		−0.617		0.109		−1.377	
	(−0.13)		(−0.25)		(0.02)		(−0.28)	
Exchange rate stability $(t-1)$	0.307		−0.063		−1.060		−1.734	
	(0.27)		(−0.06)		(−0.45)		(−0.78)	
Monetary independence $(t-1)$	−0.012		−0.840		−0.925		−2.492	
	(−0.01)		(−0.50)		(−0.26)		(−0.75)	
Δ CA† $(t-1)$	0.563***	0.467	0.455***	0.433	1.041***	0.897	0.997***	0.938
	(4.04)		(3.64)		(3.55)		(4.07)	
Primary balance† $(t-1)$	0.118		0.106		0.219		0.132	
	(1.18)		(1.15)		(1.05)		(0.73)	
(Net) public debt† $(t-1)$	0.021**	0.002	0.023**		0.048***	0.004	0.043**	
	(2.65)		(2.59)		(2.88)		(2.45)	
Δ Credit/GDP† $(t-1)$	0.004		−0.007		−0.053	−0.028	−0.060	−0.022
	(0.15)		(−0.25)		(−0.86)		(−1.04)	
Economic Freedom $(t-1)$	0.175		0.129		0.047		−0.064	
	(1.26)		(0.95)		(0.16)		(−0.24)	
Ease of Doing Business index $(t-1)$	−0.005		−0.027		−0.113		−0.125	
	(−0.13)		(−0.66)		(−1.27)		(−1.56)	
Employment protection (dismissals) $(t-1)$	1.016**		0.654		0.617		−0.020	
	(2.14)		(1.45)		(0.62)		(−0.02)	
Employment protection (temporary) $(t-1)$	−0.155		0.143		0.008		0.370	
	(−0.29)		(0.26)		(0.01)		(0.34)	
Constant	−14.039	−1.052	−9.259	−0.952	−2.782	−1.095	10.405	−0.845
	(−1.02)		(−0.67)		(−0.10)		(0.38)	
5-years FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	50	50	50	50	50	50	50	50
R^2	0.75	0.46	0.76	0.48	0.79	0.50	0.82	0.53

Notes: Numbers in parenthesis are t-statistics. ***, **, and * correspond to significance levels of 1%, 5%, and 10%, respectively. 5-years FE group the recessions into the periods 1990–94, 1995–99, 2000–04, 2004–09, 2010–14. The LASSO tuning parameter is obtained by leave-one-out cross-validation. We indicate by $t-1$ that we use values of the variables from the year before the first recession year. The effective sample of recessions is limited by the observations for which we have current-vintage data. Variables which we include from the current/final vintage are marked with a †. We take “current-vintage” values $(t-1)$ from the autumn vintage in year $t-1$.

2.A.4 The Recessions in Japan in 1997 and in Germany in 2001

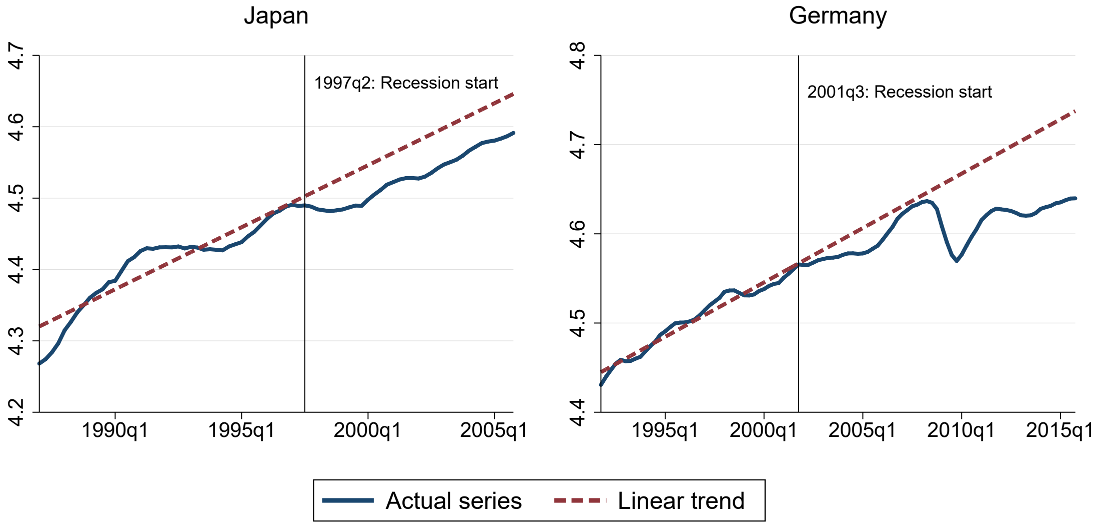

Figure 2.10. Labor productivity (in logs)

Notes: Labor productivity is measured at quarterly frequency and smoothed by taking a moving average of 4 quarters. The linear trends are calculated based on the 40 quarters before the recession start.

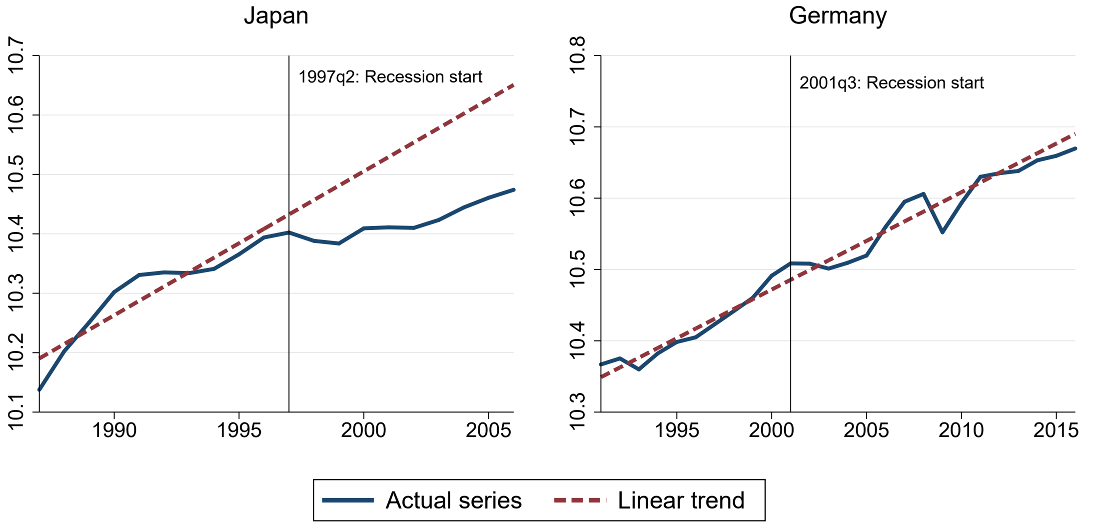

Figure 2.11. Real GDP per capita (in logs, USD PPP)

Notes: Real GDP per capita is measured at yearly frequency. The linear trends are calculated based on the 10 years before the recession start.

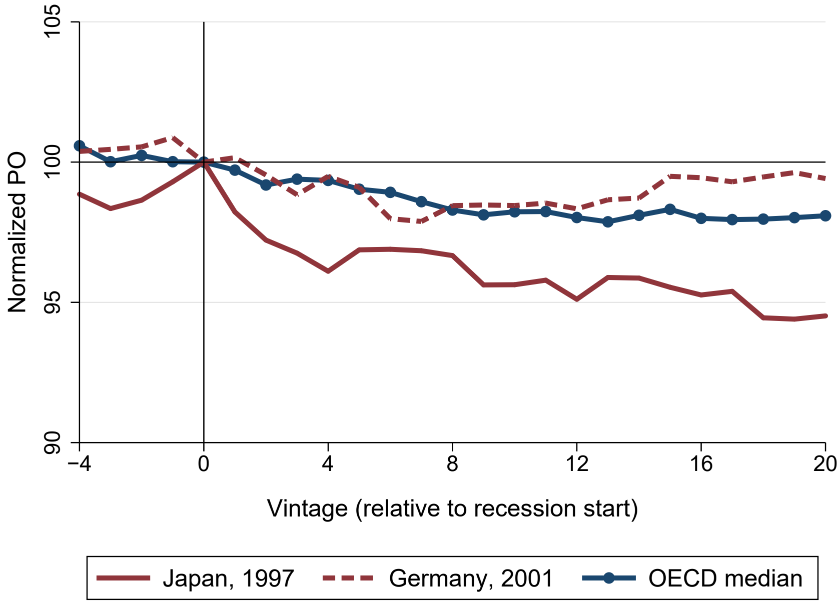

Figure 2.12. Revisions to PO estimates after recessions in Japan and Germany

Notes: The lines show the revisions to OECD PO estimates for the first year after the recession start in Japan in 1997 and in Germany in 2001 across different vintages. The data are normalized such that the value as estimated in the first vintage following the start of the recession is equal to 100. For comparison, we show the median revisions across all recessions in our sample (from Figure 2.2).

3

How Economic Crises Damage Potential Output – Evidence from the Great Recession[1]

Joint with Jonas Dovern

3.1 Introduction

The Great Recession of 2008/09 had long lasting effects on most major economies. In particular, estimates of potential output (PO) for many countries are substantially lower now than the corresponding pre-recession projections. It is well established that this is a common observation in the aftermath of major economic crises (Cerra and Saxena, 2008) and also that after 2008/09 many EU countries were among those countries which suffered most heavily following the Great Recession (Ball, 2014). It is less clear, in contrast, how large the relative importance of revisions to different components of PO has been for the overall decline of PO levels.[2] This, however, is of importance for identifying the mechanisms via which deep recessions lead to long-run economic damage and for identifying suitable policies that might help pushing economies back to their previous growth paths.

This chapter answers the questions of how much, how fast, and how persistently estimates of the capital stock, of trend labor, and of trend total factor productivity (TFP) are revised downwards after major economic crises. We consider the Great Recession and the European sovereign debt crisis as two prominent examples of major economic crises. Our newly compiled real-time data set contains data for EU member states. Our research focus differs from other related papers that only look at overall revisions of PO (or gross domestic product (GDP) for that matter) and do not look at the components (e.g., Reinhart and Rogoff, 2009, 2014; Ball, 2014; Hosseinkouchack and Wolters, 2013; Blanchard et al., 2015). Our approach also differs

[1] This chapter is based on our article in the *Journal of Macroeconomics* (Dovern and Zuber, 2020b). It is available at https://doi.org/10.1016/j.jmacro.2020.103239.

[2] We use the term "revisions" to refer to the overall change of PO estimates between different data vintages. These changes can be due to measurement error, data revisions, new data, and changes in methodology. The attempt to determine the contribution of each of these effects for overall revisions is beyond the scope of this chapter.

because most previous papers consider PO growth rates and derive revisions to the level of PO without taking revisions to pre-recession potential growth rates into account (e.g., Benati, 2012; Haltmaier, 2012; Martin et al., 2015; Coibion et al., 2018) and/or—with the exception of Furceri and Mourougane (2012) and Dovern and Zuber (2020a)—do not carefully look at the timing of PO revisions due to the lack of real-time data.

The chapter shows that revisions to different components of PO contributed equally to the substantial overall decline in estimated PO levels after the Great Recession. We are the first who systematically analyze the contributions of the revisions of components to the overall PO revision for a comprehensive real-time data set and a large country sample. By tracking revisions in real time, we provide information on the relationship of component revisions across time and on the dynamics of PO revisions made by experts over time. Our work contributes to a limited number of studies that focus on revisions to individual PO components after economic crises. Using ex-post data, Furceri and Mourougane (2012) show for 30 countries in the Organization for Economic Co-operation and Development (OECD) between 1960 and 2008 that the reductions of capital stock estimates account for most of the permanent decrease of PO after financial crises. Similarly, Haltmaier (2012) provides evidence that the capital-output ratio is the single most important contributor to lower trend output per capita after recessions for a sample of ten OECD countries while declines in trend employment and participation rates contributed to a fall of trend output only in some of the countries. With special focus on the Great Recession in the United States, Hall (2014) documents that a deterioration of the capital stock and lower TFP were the main drivers of permanently lowered trend GDP. Fernald (2015) confirms this finding over a larger sample period and argues that PO returned to a lower path after the exceptionally high TFP growth during the 2000s ended.

We find that the revision process after the European sovereign debt crisis was totally different. The European Commission (EC) seems to have fully incorporated lower PO paths in its estimates by 2011 and did not substantially lower its estimates further afterwards.

A methodological contribution of the chapter is that we conduct the empirical analysis using a comprehensive real-time data set of estimates for PO levels by the EC. Our data cover 27 real-time vintages from autumn 2005 to autumn 2018 and 27 EU countries. The data set contains information about the contribution of the capital stock, trend labor, and trend TFP, respectively, to PO growth that allows us to derive the trend levels of these components for those vintages in which component levels are not readily available. This allows us to focus on revisions to the *levels* of PO and its components. This is important because PO is often revised also for past time periods so that looking only at revisions of PO *growth*, as it has been done in most papers in this literature (e.g., Benati, 2012; Furceri and Mourougane, 2012; Martin et al., 2015; Coibion et al., 2018), systematically underestimates the true size of PO revisions.

PO is a measure that is commonly used to quantify the long-run production capacities of an economy (Okun, 1962; Havik et al., 2014). In modern macroeconomics, it is an important factor when deciding on monetary or fiscal policy. Central banks, for instance, evaluate how far the PO level is below (above) GDP to infer future inflationary (disinflationary) pressure when deciding on the appropriate monetary policy (e.g., Draghi, 2017; Yellen, 2017). Fiscal policy and, in particular, commonly used fiscal rules depend on the PO level. For example, government spending that increases productivity "raises [...] thereby future potential output, which increases fiscal space today" (Draghi, 2019). Thus, the PO level is an important indicator for designing and evaluating appropriate macroeconomic stabilization policies.

We show how sensitive PO estimates are in response to economic crises, providing evidence that policy makers should be careful to use pre-recession PO estimates when deciding on appropriate (monetary and fiscal) stabilization policies. Measurement error can lead to poor policy decisions. To mitigate the policy errors induced by a strong reliance on current PO estimation methods, policy makers could focus on observables (e.g., inflation and wage dynamics or survey data on capacity utilization), account for additional variables like the evolution of the current account or the credit growth when estimating PO (as suggested by Dovern and Zuber, 2020a), and/or account for model uncertainty by building a range of estimates (as, for example, recently done by González-Astudillo, 2019a,b). In any case, PO and the output gap remain reasonable complementary measures for decision makers (e.g., Edge and Rudd, 2016; Guisinger et al., 2018).

Since PO is defined to proxy the sustainable long-run level of output, it should in principle be independent of cyclical (temporary) fluctuations. However, it reacts to supply shocks that lead to a reassessment of trend TFP. Moreover, PO revisions could be driven by demand shocks in the presence of hysteresis (e.g., Blanchard and Summers, 1986, 1987; Lindbeck and Snower, 1986; Stadler, 1986, 1990). Such hysteresis effects can potentially work through the different components of PO. If the hysteresis effects caused structural changes in the labor market, we would expect revisions of trend labor. If they worked through a depreciation of the capital stock or a decline of investment in physical capital, we would expect estimates of the capital stock to be revised. Finally, if they caused a decline in R&D investment, we would expect revisions of trend TFP. Thus, it is of great interest to investigate which components of PO experts revise downwards after economic crises. A third explanation for PO revisions is that previous estimation errors are simply corrected. Our results on the importance of the revisions to those three components of PO address the fact that empirically little is known about the relevance of the three channels.

The remainder of this chapter is structured as follows. We present our data and explain the methodology that lies behind the PO estimates published by the EC in Section 3.2. Section 3.3 contains our empirical results. We start by analyzing the revisions of PO estimates and the contributions of the component revisions to these overall PO revisions. We then investigate the relative timing of component revisions and the cross-country differences of revisions. Further, we compare PO revisions after the Great Recession with PO revisions after the European sovereign debt crisis. Section 3.4 concludes.

3.2 Data

In this section, we decribe our data set and explain how we normalize PO and construct levels of PO components.

3.2.1 Real-Time Vintages

Our main real-time data are from two sources published by the EC in spring ("I") and autumn ("II") of each year. First, the European Economic Forecast (EEF) contains estimates of the PO level, the estimated contributions of the PO components (the capital stock, trend labor, and trend TFP) to PO growth, the estimated contributions of trend labor components (the actual working-age population, the trend labor force participation rate, the non-accelerating wage rate of unemployment (NAWRU), and trend average hours worked per head) to trend labor growth, and—since vintage 2014:I—estimates of levels of PO components. All variables are published along with four-years-ahead to five-years-ahead forecasts. Second, the Annual Macro-economic Database (AMECO) contains macroeconomic time series (such as the level of the working-age population between 15 and 64 years) along with forecasts for one year ahead. Our sample contains 27 data vintages ranging from autumn 2005 to autumn 2018. It covers the EU27 countries.[3] The vintages contain annual data starting in 1965 or in some cases later (e.g., the time series for Germany start in 1991 due to the reunification and data for new member states start at different points in the 1990s).

The use of the EC data is appealing for the following three reasons. First, it allows us not only to track PO revisions across different vintages but also to decompose these revisions into revisions of estimates of the capital stock, trend labor (which we can further decompose into detailed labor market components), and trend TFP. Second, the data allow us to analyze revisions during the Great Recession for all EU member countries. Third, the data are comparable across countries since the EC uses a production function approach to estimate PO that differs only slightly in terms of parameterization across the member countries.

---

[3] The EU27 refers to the EU member states before Croatia's accession. These countries are Austria, Belgium, Bulgaria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom. The EC stresses in a paper on the production function methodology that "equal treatment for all of the EU's Member States needs to be strictly assured" (Havik et al., 2014, p. 6). Therefore, we include all countries to make an overall assessment of how the EC's PO estimates are revised.

The EC assumes that PO is generated by a standard Cobb-Douglas production function

$$\overline{Y}_{i,t} = \overline{A}_{i,t}\,\overline{L}_{i,t}^{\alpha}\,K_{i,t}^{1-\alpha} \tag{3.1}$$

with a common labor share $\alpha = 0.65$ for all countries i and years t. Bars above variables denote the trend of the variable. PO is composed of three components. Trend TFP, $\overline{A}$, is estimated using a state-space model that decomposes TFP into a trend and a cycle and relates the latter to the degree of capacity utilization. The capital stock, K, is the unfiltered capital stock which is calculated from consumption of fixed capital. Trend labor, $\overline{L}$, corresponds to the trend of total hours worked and is calculated as the product of trend employment (which is the product of the actual working-age population, N, the trend labor force participation rate, $\overline{PR}$, and 1 minus the NAWRU, $\overline{UR}$) and the trend average hours worked per head, $\overline{H}$:

$$\overline{L}_{i,t} = N_{i,t}\,\overline{PR}_{i,t}\,(1 - \overline{UR}_{i,t})\,\overline{H}_{i,t}\,. \tag{3.2}$$

The EC applies a Hodrick-Prescott (HP) filter to the labor force participation rate and the average hours worked in order to extract the trend components. The NAWRU is obtained from a state-space model that decomposes the unemployment rate into a trend and a cyclical component making use of the Phillips curve relationship.

3.2.2 Normalization of Potential Output Data

Changes in the unit of measurement (e.g., due to the adoption of the euro) and/or base years require that we normalize PO to make it comparable across vintages and countries. Figure 3.1 shows the raw PO level estimates for Italy from all available vintages. Pictures for other countries contain similar breaks between the vintages.

We normalize PO by the deviation of PO for the year $t_0 = 2000$ in vintage v from PO for t_0 in the first available vintage v_{min} for country i:

$$\overline{Y}_{i,t}^{v} = \widetilde{Y}_{i,t}^{v} \cdot \widetilde{Y}_{i,t_0}^{v_{min}(i)} / \widetilde{Y}_{i,t_0}^{v}\,. \tag{3.3}$$

We obtain the raw level of PO, $\widetilde{Y}$, directly from the EEF vintages. The first available vintage v_{min} is 2005:II except for Bulgaria and Romania for which estimates are not reported before 2007:I. We choose $t_0 = 2000$ since it is the earliest common year for which we can normalize PO for all countries and almost all vintages.[4] This implicitly assumes that the level of PO for the year 2000 is no longer revised after 2005 but rather changes only due to changes in the base year or units of measurement.

[4] In the vintages 2014:I and 2014:II, time series of Bulgaria and Romania do not start before 2003. We therefore do not use these two vintages for the two countries.

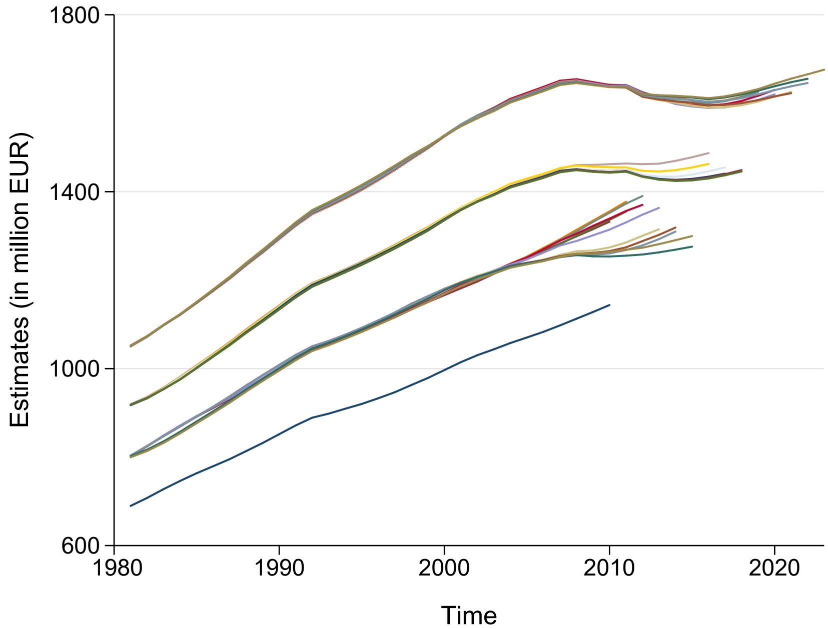

Figure 3.1. Raw data vintages of PO estimates for Italy

Notes: The plot shows estimates of PO for Italy from different EC vintages (subject to base year changes). Each line represents a time series from one particular vintage.

3.2.3 Construction of Levels of Potential Output Components

Unfortunately, information about the estimated level of the components of PO is not available in vintages from before 2014:I. Therefore, we need to construct "synthetic" levels of components. For the capital stock, trend labor, trend average hours worked per head, and the actual working-age population, we do this by resorting to information provided about the contributions of those variables to PO growth. For $X = \{K, \overline{L}, \overline{H}, N\}$, we compute the level as follows:

$$X^{v}_{i,t+1} = (1 + \dot{x}^{v}_{i,t+1})\, X^{v}_{i,t}\,, \tag{3.4}$$

where $t > t_0$, X^{v}_{i,t_0} is fixed, and $\dot{x}$ denotes the annual growth rate of the component X as implied by the growth contribution provided by the EC.[5] We set $X^{v}_{i,t_0} = 100$ for all countries i and vintages v and choose a reference year $t_0 = 2000$ since time series of growth rates do not start earlier for some countries in our sample. By using this approach, we assume that estimates of PO levels for the year 2000 do no longer change in the data vintages that we use. Additionally, this implicitly normalizes the synthetic variables.

We obtain the levels of the remaining components as follows. First, we calculate trend TFP as a residual from equation (3.1) using the normalized level of PO and the synthetic levels

[5] For instance, we multiply the growth contribution of trend labor by one divided by the labor share to get the annual growth rates.

of the capital stock and trend labor. Second, we use the synthetic levels of trend labor, the actual working-age population, and the trend average hours worked per head together with estimates of the NAWRU (which we obtain directly from the EEF) to calculate the trend labor force participation rate as a residual according to equation (3.2).

Our choice to compute the trend labor force participation rate as a residual is not arbitrary. Since vintage 2013:I, the EC has used the working-age population between 15 and 74 years. Until 2012:II, the age bracket had been 15 to 64 years. Without adjustment this would imply a systematic upward revision of the working-age population and a downward revision of the trend labor force participation rate. We avoid this methodological revision by using the working-age population between 15 and 64 years for all vintages. For vintages before 2013:I, we obtain the growth rates directly from the EEF. From 2013:I onwards, we resort to the growth rates implied by the level series provided by AMECO.

3.3 Empirical Results

Revisions to PO and PO components may vary along different dimensions. In the first two subsections, we analyze the revisions to PO and PO components after the Great Recession. Thereafter, we explore two dimensions in more detail: the timing of component revisions and how revisions differ across countries. Finally, we compare the revisions to PO after the Great Recession with the revisions to PO after the European sovereign debt crisis.

3.3.1 Revisions of Potential Output Levels

We begin our analysis by tracking how the EC revises its PO estimates for specific years following the Great Recession. To make estimates comparable across countries, we normalize them such that the PO estimate in year t and vintage 2008:I is 100 for each country. The choice of the vintage 2008:I as the first recession vintage is driven by the data. 14 EU economies are in a recession in the second quarter of 2008 and six additional economies slip into a recession in 2008Q3. Hence, we will refer to 2008 as the first recession year for all countries.[6] Figure 3.2 shows the median revisions of PO estimates for the years 2008 and 2012. By focusing on the median of revisions, we limit the effect that idiosyncratic and eruptive revisions have on our results. We believe that the rather stable development of the median revisions across time is an indication that this strategy is successful.

Three characteristics stand out. First, the median revisions are substantial and increase the longer the forecast horizon is. The PO estimate for 2008 is, on average, revised by more than

[6] We use the simple method of Harding and Pagan (2002) to identify recessions based on quarterly real GDP from the latest available vintage. The algorithm searches for business cycle peaks and troughs requiring a cycle length of five quarters and a length of each business cycle phase of at least two quarters. Table 3.3 in the appendix summarizes the identified recessions.

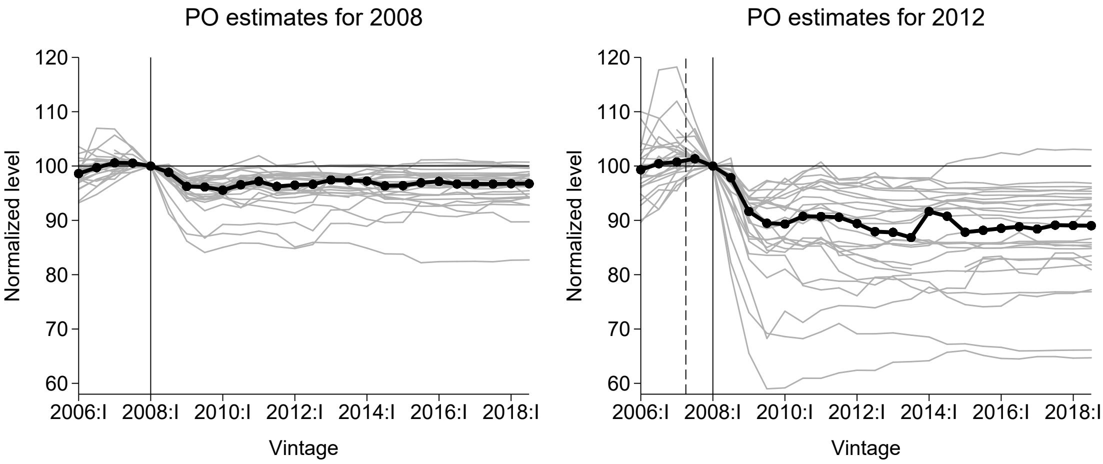

Figure 3.2. Revisions of PO estimates after the Great Recession

Notes: The plots show the median revisions of the EC's PO estimates for 2008 and 2012 across different vintages. Grey lines represent single country revisions. Levels are normalized such that they are 100 in 2008:I for every country and year. Values to the left of the dashed line are extrapolated as in Chapter 2.

−3% in the long run. For 2012, the median PO estimate for 2018:II is revised by −11%.[7] Second, most of the median revisions of PO are made within one year (equivalent to two vintages). PO estimates remain persistently below the pre-crisis estimates. Fluctuations afterwards are country-specific and small.[8] Third, revisions are very heterogeneous in magnitude across countries, but positive revisions are scarce. The standard deviation of the revisions of PO estimates between 2008:I and 2018:II nearly triples from 3.7 for 2008 to 9.0 for 2012. Downward revisions of PO estimates are substantial for small EU economies while many large economies observe moderate revisions between 2008:I and 2018:II (e.g., the revision of the PO estimate for 2008 is +0.1% for Germany, −2.3% for France, and −3.3% for Italy).

To evaluate the size of revisions, we need to consider alternative PO estimates. First, we use estimates from a purely statistical application. We obtain these estimates by applying the one-sided HP filter (with $\lambda = 100$) to annual time series of real GDP (and additional GDP forecasts) in each vintage. The use of the one-sided HP filter is appealing as it is purely backward looking in contrast to the two-sided HP filter. Thus, only data revisions for previous years but not future realizations can change the trend today. Second, we use PO estimates from the OECD Economic Outlook. These estimates are especially interesting because the OECD uses a production function approach to estimate PO as well. We normalize both sets

[7] Based on a sample of OECD estimates of PO for 95 recessions, Dovern and Zuber (2020a) find that estimates are, on average, revised by −1.9% for the first year after a recession and −4.1% for the fifth year after a recession during the first ten years after a recession. This comparison shows that the Great Recession was a very severe crisis that caused more long-run economic damage than an "average recession".

[8] The median revision for 2012 temporarily jumps in 2014:I and 2014:II due to the missing estimates for Bulgaria and Romania.

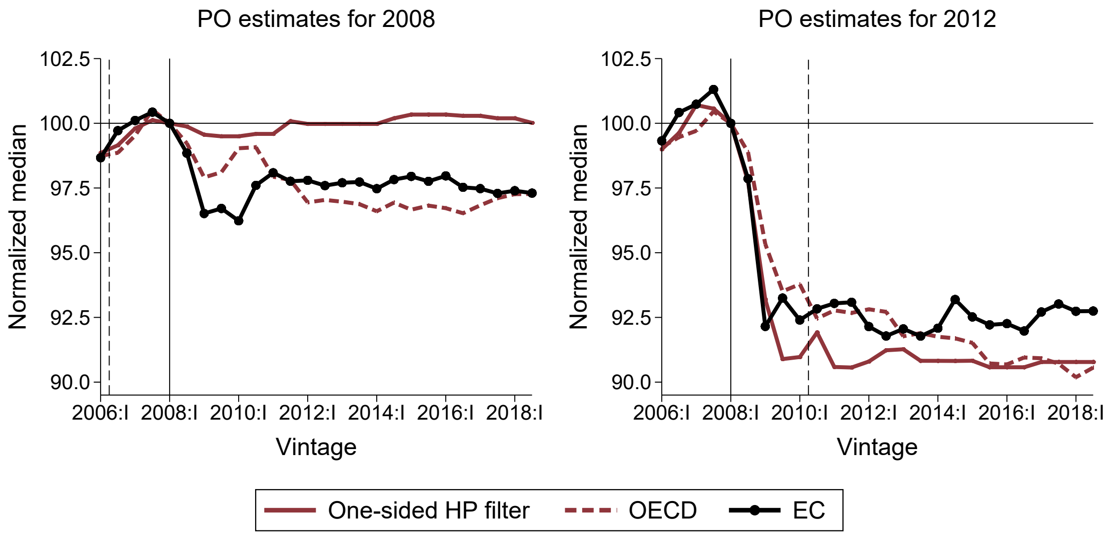

Figure 3.3. Comparison of PO revisions from the EC, OECD, and the HP filter

Notes: The plots show the median revisions of PO estimates obtained from the one-sided HP filter for 2008 and 2012 across different vintages. For comparison, we show the median PO revisions of the OECD and EC. Medians are normalized such that they are 100 in 2008:I for every year. Estimates of the OECD and the HP filter which are to the left of the dashed line are extrapolated as in Chapter 2. We omit the extrapolation line of EC estimates (see Figure 3.2 for its location). The medians are based on the 17 countries that are covered by both institutions.

of PO estimates in the same way as we normalize the EC's PO estimates. Unfortunately, the OECD only provides PO estimates for a subsample of European countries. To make the different estimates comparable, we concentrate on those 17 countries which are covered by both the EC and the OECD.[9]

Two points are noteworthy in Figure 3.3. First, the EC revises the PO estimates much more quickly than the OECD. This is especially evident for 2012. While the EC's PO estimate for 2012 is hardly revised after vintage 2009:I (revisions range from −6.75% to −8.25%), the OECD's PO estimate for the same year is continuously revised downwards (from around −5% in 2009:I to nearly −10% in 2018:II). Second, the EC median revisions are similar or more moderate compared to the OECD revisions in the long run. For 2008, the revision is roughly −3% until 2018:II for the estimates of the EC and OECD. Since the one-sided HP filter is purely backward looking, we do not see any revision for 2008. For 2012, both alternative estimates are about two percentage points lower than the EC median in 2018:II. We conclude that the EC's PO estimates are the least revised estimates for 2012 in the set of available alternatives. The gradual revisions of the OECD indicate that the OECD attributes a larger part of the PO estimate to cyclical factors than the EC does.

[9] The EU countries covered by both institutions are Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Poland, Portugal, Spain, Sweden, and the United Kingdom.

Large revisions of PO estimates can lead to substantial distortions. A simple thought experiment quantifies the distortionary effect of PO revisions on fiscal space. Fiscal space generally describes the government's room for budget adjustments (increasing government spending / lowering taxes) while preserving sound and sustainable public finances or respecting prevailing fiscal rules. A very simple measure of fiscal space in our context is the difference between the structural balance and the admissible minimum structural balance of –3% that is defined in the Maastricht Treaty. We consider Italy as example since Italy's PO revision for the year 2008 between 2008:I and 2018:II is close to the median. Based on PO estimates from 2018:II, the structural balance for Italy for 2008 is −3.6%. This implies that there is no fiscal space at all. In contrast, given the PO estimate and the corresponding output gap as of 2008:I, one obtains a structural balance of only −1.9% (European Commission, 2008, p. 79). So without considering the post-recession revisions to PO, the estimate of fiscal space for that year is 1.1% of PO, substantially larger than zero.

3.3.2 Revisions of Potential Output Components

The data allow us to break down PO revisions into revisions of the capital stock, trend labor, and trend TFP. Figure 3.4 shows the median revisions of the three PO components after the Great Recession. As for PO before, we apply a normalization such that the values in vintage 2008:I for the shown years are equal to 100.

The median component estimates stay persistently below the pre-crisis value in five of the six cases. For 2008, median revisions of trend labor are largest (−1.8%) followed by small revisions of the capital stock (−0.8%), with no visible revisions of trend TFP. For 2012, the median revisions are more balanced between the components, ranging from −2.5% (capital stock) to −3.4% (trend TFP) for the most recent vintage. The median revisions show further that the revision of trend labor is gradual especially for the year 2012. The EC seems to continuously receive new information on the labor market and updates its trend labor estimates accordingly to lower values. In contrast, the revision of the capital stock is a one-off revision within one year after the recession started. Between 2008:II and 2009:I, the estimate of the capital stock for 2012 is revised by −2% which is already more than three quarters of the long-run capital stock revision. The capital stock is estimated and projected using forecasts of the investment ratio. Sudden revisions of the investment ratio translate immediately into revisions of the capital stock as the EC does not filter the latter. Therefore, we do not observe gradual adjustments to the capital stock revisions.[10] For 2012, the median investment ratio is, for instance, revised by −2.1 percentage points between 2008:II and 2009:I.

[10] If the capital stock was estimated by an HP filter, the weights for the end-of-sample years would continuously change with each data vintage that contains revised or new forecasts.

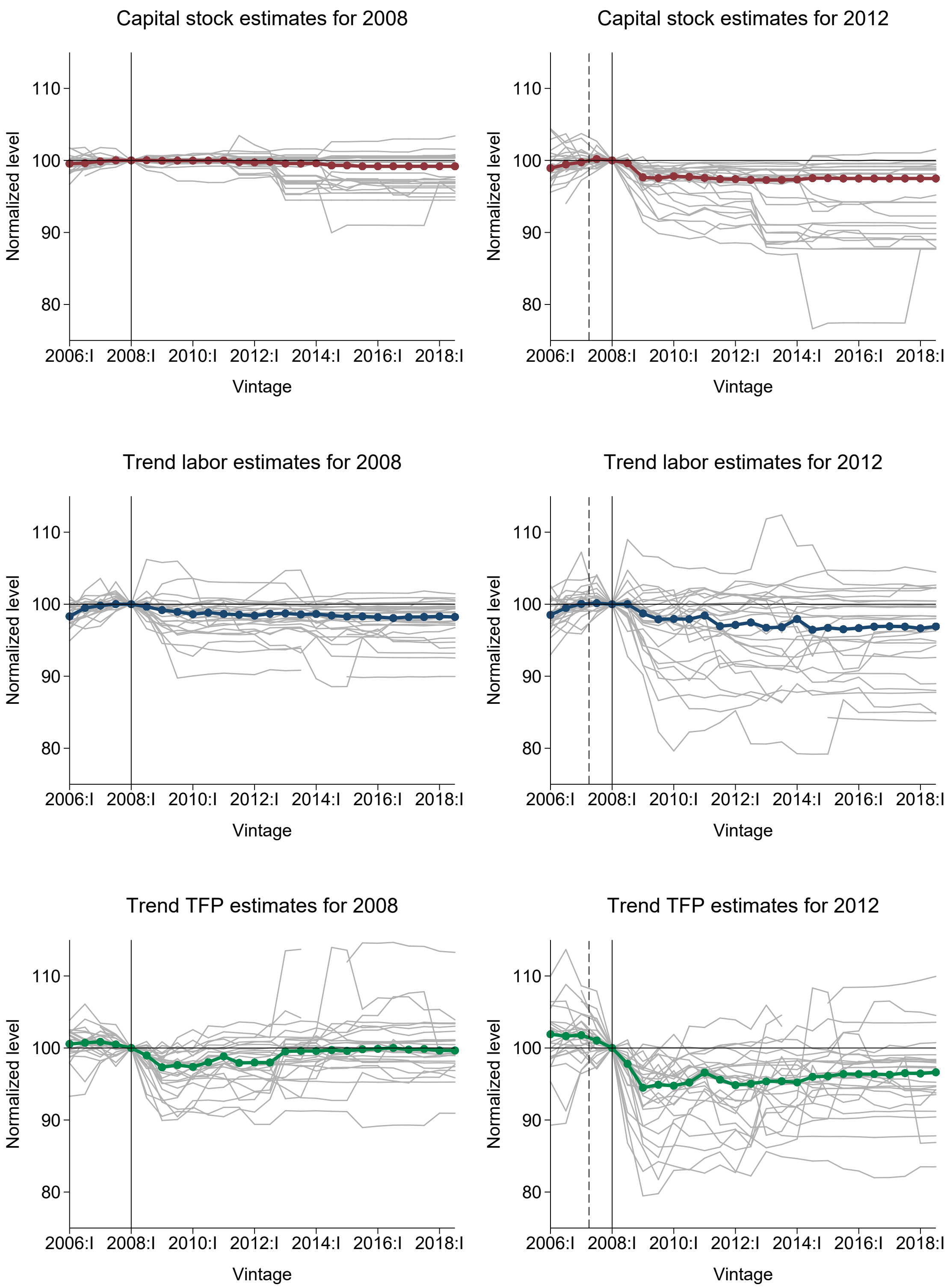

Figure 3.4. Revisions of estimates of PO components after the Great Recession

Notes: The plots show the median revisions of the EC's component estimates for 2008 and 2012 across different vintages. Grey lines represent single country revisions. Levels are normalized such that they are 100 in 2008:I for every country and year. Values to the left of the dashed line are extrapolated as in Chapter 2. For 2014:I and 2014:II, estimates of trend labor and trend TFP are missing for Bulgaria and Romania.

To measure how much, on average, the revisions of a component contribute to the overall PO revisions, we calculate the deviation of the level of variable X in vintage v_k from the level in vintage $v_0 = 2008{:}\mathrm{I}$ for the years $t = \{2008, 2012\}$. Specifically, we calculate for $X = \{\overline{Y}, K, \overline{L}, \overline{A}\}$ the log difference $\Delta x_{i,t}^{v_0 \to v_k} = \log(X_{i,t}^{v_k}) - \log(X_{i,t}^{v_0})$. By doing so, we can decompose PO revisions into revisions of the three components:

$$\Delta \overline{y}_{i,t}^{v_0 \to v_k} = \Delta \overline{a}_{i,t}^{v_0 \to v_k} + \alpha \Delta \overline{\ell}_{i,t}^{v_0 \to v_k} + (1-\alpha) \Delta k_{i,t}^{v_0 \to v_k} \,. \tag{3.5}$$

Table 3.1 shows the mean revisions of PO and the contributions of PO components for the years 2008 and 2012 across different vintages. We find that trend TFP revisions dominate in the short run while in the long run revisions of trend labor and the capital stock are more prominent. In 2009:I, revisions of trend TFP account for 74% (60%) of the overall PO revision for 2008 (2012). Compared to the most recent vintage, revisions of trend labor (59%) account for the most of the overall PO revision for 2008 and revisions of the capital stock (39%) account for the most for 2012. The strong reaction of trend TFP in the short run suggests that trend TFP revisions may lead revisions of the capital stock and/or trend labor. Experts initially attribute the PO revisions to negative technology shocks but revise this view by putting more emphasis on labor market changes and a reduced capital stock. We analyze the relative timing of component revisions in more detail in the next section.

A closer look at the contribution of subcomponent revisions to the overall revision of trend labor reveals that revisions of the NAWRU primarily contribute to the overall trend labor revisions (see Appendix 3.A.2). The NAWRU revisions account for 56% of the overall revisions of trend labor in the most recent vintage for 2008 and 2012. In particular, the average NAWRU estimate for 2012 increased by 3.4 percentage points from 5.6% in 2008:I to 9.0% in 2018:II. This corresponds quite well to the existing literature that shows that NAWRU estimates are very cyclical (e.g., Lendvai et al., 2015; Heimberger et al., 2017). Similarly, Oulton and Sebastiá-Barriel (2017) provide evidence that banking crises substantially reduce the long-run level of employment which also goes along with a higher unemployment rate.

Although the EC changed the calculation of the components over the vintages, these changes do not systematically drive our results. In our sample, we identify four major changes in the methodology.[11] First, since 2010:II, the EC has used a Kalman filter to estimate trend TFP. Before, the EC had used an HP filter approach which was more revision prone and returned less realistic estimates in the short and medium run. Second, the EC switched in 2013:I from measuring the capital stock by the perpetual inventory method to using the capital stock series by AMECO, which builds on data on the consumption of fixed capital.

[11] We identify methodological changes from Denis et al. (2006), D'Auria et al. (2010), and Havik et al. (2014) who document the EC approach to estimate PO and the changes to it.

Table 3.1—Revisions of estimates after the Great Recession

	Change vs. vintage 2008:I (in %)					
	2006:I	2007:I	2009:I	2010:I	2013:I	2018:II
Estimates for 2008						
PO	−1.61	0.53	−4.81	−5.03	−3.92	−4.05
Capital stock	−0.44	−0.13	−0.10	−0.07	−1.39	−1.57
Trend labor	−1.54	−0.14	−1.17	−1.88	−1.76	−2.37
Trend TFP	0.37	0.80	−3.54	−3.08	−0.77	−0.11
Estimates for 2012						
PO	−1.11†	1.93†	−12.09	−13.61	−14.67	−13.95
Capital stock	−0.67†	−0.14†	−2.68	−3.29	−5.30	−5.39
Trend labor	−1.46†	0.24†	−2.18	−3.85	−4.37	−4.37
Trend TFP	1.02†	1.83†	−7.23	−6.47	−5.00	−4.19

Notes: The table reports full sample mean revisions of PO and the three PO components. † indicates that the computation of the result involves our extrapolation of estimates as in Chapter 2.

Especially for newer member states with a limited history of data points, this leads to large changes. Third, since 2014:I, the EC has used a non-centered NAWRU for three quarters of the countries and has changed how it extends the NAWRU forecasts. Fourth, also since 2014:I, the European System of National and Regional Accounts (ESA) 2010 replaced the previous accounting standard ESA 95. Carefully inspecting Figure 3.4, we find that the medians in the vintages before the methodological changes are roughly the same as the medians in the vintages shortly after the methodological change. A closer look at the single revisions of each country at the three time points reveals that estimates of trend labor or trend TFP are not systematically revised upwards or downwards in 2010:II or 2014:I. For the capital stock, we observe downward revisions for some countries from 2012:II to 2013:I (and to a lesser extent from 2013:II to 2014:I), but they do not affect the median.[12]

3.3.3 Timing of Component Revisions

We investigate the relative timing of the revisions by looking at the correlations between revisions to the PO components. We calculate a revision time series for each PO component $X = \{K, \overline{L}, \overline{A}\}$ and country i according to $\Delta x_{i,t}(v_k) = \log(X_{i,t}^{v_k}) - \log(X_{i,t}^{v_{k-1}})$ where year t is the maximum year available in both vintages.

Figure 3.5 shows the distributions of pairwise correlations between those time series across countries. Three correlations stand out. First, revisions of trend labor and the capital stock contemporaneously tend to move in the same direction. The median correlation is 23%. In

[12] We cannot exclude the possibility that the upward revisions of trend TFP for 2008 from 2012:II to 2013:I come from the change in the capital stock measurement. For the following years up to 2012, these revisions, however, do not affect the median.

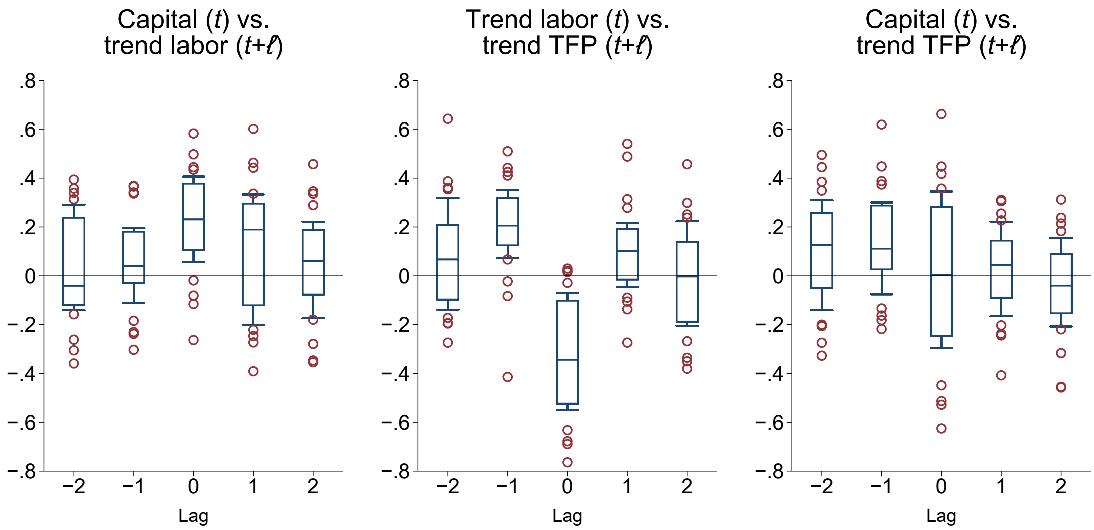

Figure 3.5. Pairwise correlations of revisions of PO components

Notes: The plots show the pairwise correlation between the PO components for different lags ℓ. Boxes show the interquartile range, the interior line the median, and whiskers the 17$^{\text{th}}$ and 83$^{\text{rd}}$ percentiles. Values below the 17$^{\text{th}}$ percentile or above the 83$^{\text{rd}}$ percentile are depicted by circles.

a boom, higher labor demand leads to a downward revision of the NAWRU and therefore to an upward revision of trend labor. At the same time, the investment ratio is revised upwards leading to a positive revision of capital. Second, the median of the contemporaneous correlation between trend labor and trend TFP is −34%. The revisions of the two components contemporaneously move in opposite directions. Only for three countries the correlations between those revision series are slightly positive. Experts initially seem to attribute changes in PO to technology shocks. Later, they revise this view by putting more weight on labor market changes. Third, trend TFP revisions lead, on average, revisions of trend labor by one vintage. The median correlation is around 21% between trend labor revisions and lagged trend TFP revisions and only three out of 27 countries exhibit a negative correlation. Thus, it seems that technology shocks initially contribute in large parts to the PO revision while later this view is revised. Qualitatively, the relation between trend labor and trend TFP holds between the capital stock and trend TFP, too.

Non-contemporaneous correlations indicate how experts adjust revisions in the future. These relationships are particularly interesting since one could infer the direction of future revisions. We explore these non-contemporaneous relationships more formally by conducting Granger causality tests as formalized by Dumitrescu and Hurlin (2012) for panel data. If the test rejects the null hypothesis, we reject that a revision series A does not Granger-cause another revision series B for each country. Hence, revision series A Granger-causes revision

Table 3.2—P-values from Granger causality tests

Granger causality from / to	Capital stock	Trend labor	Trend TFP
Capital stock	–	0.810	0.105
Trend labor	0.019	–	0.069
Trend TFP	0.170	0.501	–

Notes: The table reports p-values of the standardized test statistic of Dumitrescu and Hurlin (2012). The lag length is one. The choice is based on the Akaike criterion. We exclude the revision series for Bulgaria and Romania since the revision series start later and the Dumitrescu-Hurlin test requires a balanced panel.

series B for at least one country or more. We choose a lag length of one based on the Akaike criterion. Table 3.2 reports the p-values of the Dumitrescu-Hurlin test.

In line with our previous result, we reject the hypothesis that trend TFP revisions do not Granger-cause trend labor revisions at the 10% level. The relationship between trend TFP revisions and subsequent revisions of the capital stock is similar although the test result is less clear (with a p-value of 0.105). In addition, we reject the hypothesis that revisions of the capital stock do not Granger-cause revisions of trend labor at the 5% level. This is plausible since trend labor is revised gradually while the capital stock tends to be changed in one-off revisions. Practitioners could exploit the information presented in this section to improve their forecasts of the upcoming revisions of the capital stock and trend labor by anticipating positive revisions to the capital stock and trend labor in the following vintage when they observe positive revisions to trend TFP in the current vintage.

3.3.4 Cross-Country Differences

Revisions of component estimates are very heterogeneous across countries and positive revisions are not as rare as they are for PO estimates as Figure 3.4 shows. For 2008, approximately one third of the countries experience a positive revision of the capital stock and trend TFP if one compares the estimates from the most recent vintage to the estimates from 2008:I.

Dovern and Zuber (2020a) show that country characteristics systematically drive the revision of PO estimates after recessions. It is reasonable to expect that country characteristics can explain differences in revisions of component estimates, too. For example, one could conjecture that the EU's freedom of movement is one driver behind downward revisions of trend labor in countries that are severely affected by the recession and have weak social insurance institutions. This out-migration from severely hit countries to less affected countries has been an important concern in the EU and has been extensively covered by the media (e.g., Paul Krugman in the New York Times Magazin, 2011). To test in general if revisions of PO and its components are systematically different from certain groups of EU countries, we split

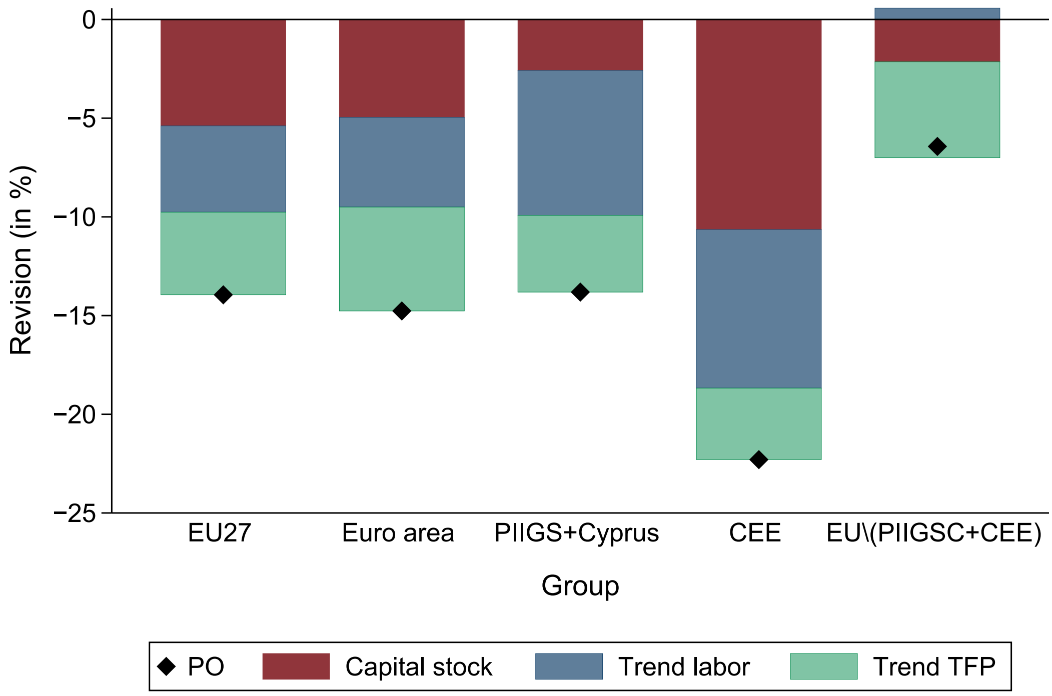

Figure 3.6. Country group revisions for 2012

Notes: The plot shows the mean contribution of component revisions to the overall PO revision between 2008:I and 2018:II for different country groups for 2012.

the EU27 countries into three groups: (a) the six countries Portugal, Ireland, Italy, Greece, Spain, and Cyprus (PIIGS+Cyprus) which suffered the most from the Great Recession, (b) the Central and Eastern European (CEE) countries[13], and (c) the remaining eleven countries which have been hit less by the Great Recession.[14] In addition, we calculate the mean revisions for the 19 euro area countries to test whether being part of the monetary union made a difference. For instance, members of the euro area could suffer more from (asymmetric) negative demand shocks in the long run since they possess less stabilization tools.

We find that revisions of PO and the PO components are significantly different for PIIGS+Cyprus and the CEE countries while we find no difference between euro area members and the remaining countries. Figure 3.6 shows the mean revisions for the different subgroups for 2012. For PIIGS+Cyprus, trend labor revisions contribute to more than a half (53%) of the overall PO revisions. Labor migration partly explains these revisions (Appendix 3.A.2). For the CEE countries, the main component behind PO revisions is the capital stock (which

[13] The ten CEE countries are Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia.

[14] Controlling for these three groups instead of specific country characteristics like the current account deficit, the private credit evolution, and other macroeconomic variables—as Dovern and Zuber (2020a) do—is a sensible choice given the small number of observations (only 25 recessions in our data set). Further analysis would ask too much from the data.

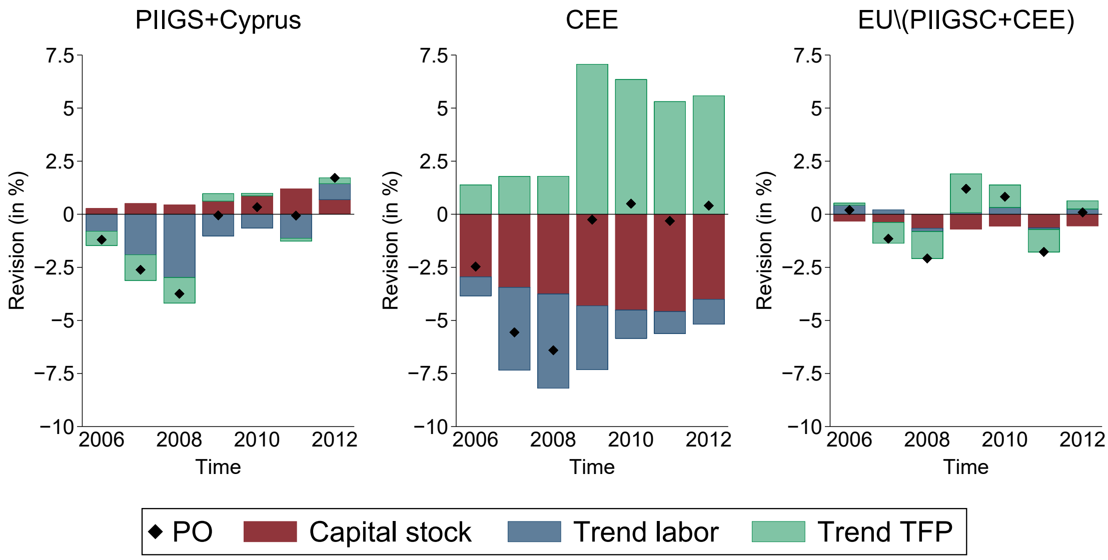

Figure 3.7. Revisions across groups and years

Notes: The plots show the mean contribution of component revisions to the overall PO revision between 2008:I and 2018:II for different country groups for 2006–2012.

contributes 48%). In the beginning of the 2000s these ten countries had very high investment ratios. These investment ratios were revised substantially downwards after 2008, which, in turn, led to the large revisions of the capital stock. The average of the pre-recession estimate of the investment ratio for 2012 was 33% for the CEE countries in 2008:I while it dropped to 22% in the latest available vintage. For comparison, the EU average dropped by only six percentage points in the same period. For countries neither being PIIGS+Cyprus nor CEE countries, trend TFP revisions account for roughly three quarters of the overall PO revision. This agrees with the notion that TFP is the major driver of growth in highly developed countries.

Revisions of component estimates for CEE countries are much larger than for other countries. We compare estimates for the years $t = \{2006, ..., 2012\}$ from the spring vintage of the same year t with estimates for year t from the most recent vintage. For $X = \{\overline{Y}, K, \overline{L}, \overline{A}\}$, we calculate these revisions by $\Delta x_{i,t}^{v_{\mathrm{I}}(t)\to v_{max}} = \log(X_{i,t}^{v_{max}}) - \log(X_{i,t}^{v_{\mathrm{I}}(t)})$ where $v_{\mathrm{I}}(t)$ refers to the spring vintage in year t and v_{max} = 2018:II. Figure 3.7 shows the results. First, we see large downward revisions of PO for the years 2006–2008. The revisions reflect the experts' changed assessment of the economic conditions after the start of the Great Recession. The relevance of component revisions differs by countries for these years.[15] Second, PO revisions after 2008 fluctuate only little within a band of ±1.8% around zero. Experts have already revised the PO estimates to incorporate the worsened economic outlook after the Great Recession. The relatively small

[15] In fact, Figure 3.7 shows that the revision of different components is particularly relevant for each country group and for all years between 2006 and 2012. It is therefore a generalization of our finding from Figure 3.6.

band width suggests that the European sovereign debt crisis in 2011–2012 did not have noticeable long-run effects on the estimates of the years before. Third, trend TFP revisions for CEE countries are more than ten times larger than the actual PO revision for the years after 2008. For the other countries, component revisions are in general smaller than the overall PO revision. Year by year, experts initially seem to overestimate the contribution of the capital stock to PO for CEE countries. Later, this culminates in large downward revisions of the capital stock and large upward revisions of trend TFP. The pattern reveals a serious overestimation of the capital stock for CEE countries in real time. The same applies to the estimates of trend labor for CEE countries, albeit to a lesser extent.

3.3.5 The Great Recession versus the European Sovereign Debt Crisis

The European sovereign debt crisis of 2011 led to double-dip recessions in several European countries. Did PO revisions during the European sovereign debt crisis evolve similarly to PO revisions during the Great Recession? We provide answers to this question by comparing the PO revisions after 2011 for 2011 (the first year of the European sovereign debt crisis) to the PO revisions after 2008 for 2008 (the start of the Great Recession) and the revisions for the four-year-ahead PO estimates (target years 2015 and 2012, respectively), too.

Figure 3.8 shows that the PO revisions after the European sovereign debt crisis differ significantly in size compared to the PO revisions after the Great Recession shown in Figure 3.2. We compare estimates for 2011 with estimates for 2008 and estimates for 2015 with estimates for 2012. The median revision of PO estimates for 2011 is nearly −1% if one compares the estimate in 2011:I with the final release in 2018:II. In contrast, the median revision of the PO estimates for 2008 was more than −3% in the long run. The same is valid for the four-year-ahead PO estimate: the median for 2015 is revised by −2.8% which is almost one quarter of the median PO revision for 2012 which we observed after 2008. Individual country estimates are again very heterogeneous. There are more positive revisions than after 2008. Similar to the pattern in the Great Recession, revisions for the large EU economies are moderate.[16] In general experts seem to have fully incorporated lower trend growth by 2011 and did not, on average, expect a further decline in trend growth due to the European sovereign debt crisis, either. The difference to the years following the Great Recession is likely to be due to two reasons. In contrast to the Great Recession, the European sovereign debt crisis mainly affected only some but not the majority of European countries. Also, its impact on average growth rates was much smaller; the median GDP growth between 2008 and 2010 was −2.6%

[16] Ireland experienced a large positive PO revision for 2015. Low corporate taxes attracted large multinational enterprises (e.g., Apple), which relocated to Ireland together with their intellectual property. Revenues from this intellectual property led to a one-off growth of GDP by 25.1% for 2015 and similarly increased PO measured in 2018:II. The OECD (2016) estimates that the GDP growth for 2015 would be just around 4.5% if one excluded the one-off effect of relocation of multinational corporations. Angerer et al. (2016) provide further details.

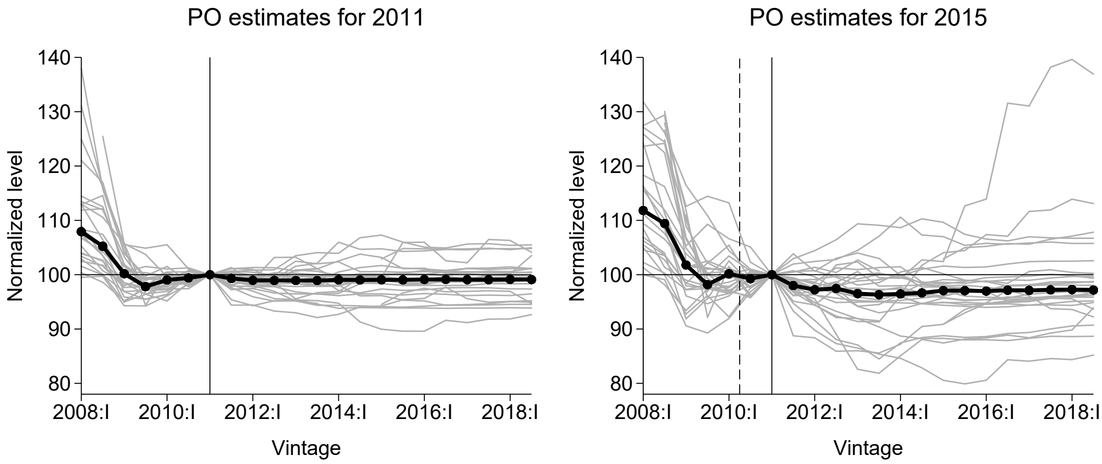

Figure 3.8. Revisions of PO estimates after the European sovereign debt crisis

Notes: The plots show the median revisions of the EC's PO estimates for 2011 and 2015 across different vintages. Grey lines represent single country revisions. Levels are normalized such that they are 100 in 2011:I for every country and year. Values to the left of the dashed line are extrapolated as in Chapter 2.

in our sample which is much less than the median GDP growth of 2.5% between 2012 and 2014.

3.4 Conclusion

Economic crises lead to downward revisions of the level of PO, but little is known about the underlying factors that drive these revisions. To fill this gap, this chapter looks at revisions by the EC to the components of PO for all EU countries after the Great Recession and the European sovereign debt crisis. We answer the questions of how much, how fast, and how persistently it has revised its estimates of the capital stock, of trend labor, and of trend TFP.

We document that estimates of PO were revised substantially downwards in response to the Great Recession and that revisions to the different components contributed about equally to this decline in the long run. The timing of revisions to the individual components is very different though. Briefly after the Great Recession, revisions of trend TFP estimates dominate due to an initial "overshooting" which is partly reversed in subsequent years. Also revisions of capital stock estimates happen very quickly after the start of the Great Recession. In contrast, it takes time before trend labor estimates are fully revised.

We also document important differences in the relevance of component revisions for the overall decline in PO estimates across countries. For the CEE countries, a strong reduction of capital stock estimates is the main source of overall PO revisions. For the group of euro area countries that suffered heavily from the crisis (PIIGS+Cyprus), the major factor behind overall PO revisions is a reduction of trend labor estimates. Technology shocks seem to be the main driver behind lower PO estimates for the remaining countries.

Our results have important policy implications. Policy makers must be aware that revisions to PO after economic crises can be huge and visible only with a delay. That implies that measures implemented based on early estimates of PO are likely to turn out to be too expansionary when viewed in light of subsequent revisions to PO estimates as we demonstrate in Section 3.3.1.[17] Our cross-country results demonstrate that different policy actions might be needed to push countries back to their previous growth trends. While stimulating investment could, for instance, help to reach a higher path for PO levels in CEE countries, the PIIGS+Cyprus countries should focus on labor market policies to push employment back to previous levels.

In a wider context, our results provide tentative evidence on the mechanisms by which recessions, or negative shocks in general, affect an economy's potential. Thus, our findings are informative for the development of macroeconomic models that explicitly model how different types of permanent *and* transitory shocks affect PO in the long run.

[17] Another question—not addressed in this chapter—is to what extent macroeconomic policies can help mitigate the long-run damaging effect of economic crises on PO documented in this chapter.

3.A Appendix

3.A.1 List of Identified Recessions

Table 3.3—Recessions which start between 2007Q2 and 2009Q1

#	Country	Start	Length	Depth
1	Austria	2008q2	5	−5.29
2	Belgium	2008q3	4	−3.81
3	Bulgaria	2009q1	4	−6.43
4	Cyprus	2008q4	4	−2.50
5	Czech Republic	2008q4	3	−5.88
6	Denmark	2008q1	6	−7.07
7	Estonia	2008q1	7	−20.89
8	Finland	2008q1	6	−10.01
9	France	2008q2	5	−3.92
10	Germany	2008q2	4	−6.94
11	Greece	2007q3	26	−27.44
12	Hungary	2008q3	7	−7.73
13	Ireland	2007q2	11	−10.90
14	Italy	2008q2	5	−7.88
15	Latvia	2007q4	12	−22.70
16	Lithuania	2008q3	6	−16.78
17	Luxembourg	2008q1	6	−7.88
18	Malta	2008q4	2	−5.11
19	Netherlands	2008q3	4	−4.36
20	Portugal	2008q2	4	−4.33
21	Romania	2008q4	2	−8.50
22	Slovenia	2008q3	4	−9.52
23	Spain	2008q3	6	−4.62
24	Sweden	2008q1	5	−7.43
25	United Kingdom	2008q2	5	−6.26

Notes: “Start” corresponds to the year and quarter in which the recession begins. “Length” refers to the duration of a recession in quarters. “Depth” refers to the maximum depth of a recession (in % of the pre-recession peak level of output). Note that the algorithm does not identify a recession for Poland and Slovakia in the time period of the Great Recession.

3.A.2 A Closer Look at the Labor Market

A nice feature of the EEF is the availability of estimates of trend labor components. Using estimates of the working-age population, the trend labor force participation rate, the NAWRU, and the trend average hours worked per head, we measure the share of each component revision in the overall revision of trend labor. Formally, we calculate revisions between vintages $v_0 = 2008\text{:I}$ and v_k for estimates of $X = \{\overline{L}, N, \overline{PR}, (1 - \overline{UR}), \overline{H}\}$ and years $t = \{2008, 2012\}$ by $\Delta x_{i,t}^{v_0 \rightarrow v_k} = \log(X_{i,t}^{v_k}) - \log(X_{i,t}^{v_0})$. By doing so, we can decompose the overall trend labor revisions into:

$$\Delta\overline{\ell}_{i,t}^{v_0 \rightarrow v_k} = \Delta n_{i,t}^{v_0 \rightarrow v_k} + \Delta\overline{pr}_{i,t}^{v_0 \rightarrow v_k} + \Delta(1 - \overline{ur}_{i,t})^{v_0 \rightarrow v_k} + \Delta\overline{h}_{i,t}^{v_0 \rightarrow v_k}. \tag{3.6}$$

Table 3.4 shows that NAWRU revisions make up the largest share in the overall revisions of the trend labor estimates. They explain 56% of the overall trend labor revisions for 2008 and 2012 between 2008:I and 2018:II. For 2012, the average NAWRU estimate increased from 5.6% in 2008:I to 9.0% in 2018:II. Countries, which were severely hit by the Great Recession, faced even larger revisions. PIIGS+Cyprus experienced an average increase of the NAWRU by 6.1 percentage points resulting in an average NAWRU estimate of 12.7% in 2018:II.

We further split the revisions of trend labor and its components into two periods: 2008:I–2013:I and 2013:I–2018:II. The split shows that the EC revises estimates of trend labor mainly in the first period and that these revisions primarily come from revisions of the NAWRU and trend average hours worked per head. For 2008, trend labor is, on average, revised by roughly three quarters of its overall revision between 2008:I and 2013:I. For 2012, the total revision of trend labor happens, on average, until 2013:I.

Finally, we investigate whether revisions of estimates of trend labor components differ across countries. Our approach is the same as in Section 3.3.4. Figure 3.9 shows that the NAWRU revision dominates in all groups. The NAWRU revision makes up 60% of the trend labor revision for PIIGS+Cyprus and 41% for the CEE countries. For the CEE countries and the EU excluding PIIGS+Cyprus and CEE countries, the revision of the working-age population is almost equally important. The revision of the working-age population contributes 37% to the trend labor revision in CEE countries and the share is larger than for all other groups. Labor migration from CEE countries to the rest of Europe may explain the large downward revision of the working-age population for CEE countries by −3% and the upward revision of the working-age population for the EU excluding PIIGS+Cyprus and CEE countries by 0.7%. Labor mobility may serve as an adjustment mechanism that reduces the output gap of individual countries. Through emigration, the unemployment rate decreases towards the NAWRU. In the immigrant-receiving country, the tightness of the labor market decreases and the unemployment rate increases towards the NAWRU.

Table 3.4—Detailed revisions of estimates of trend labor components

	2008			2012		
	2008:I–2018:II	2008:I–2013:I	2013:I–2018:II	2008:I–2018:II	2008:I–2013:I	2013:I–2018:II
Trend labor	−2.37	−1.76	−0.61	−4.37	−4.37	0.00
1 − NAWRU	−1.33	−1.05	−0.28	−2.45	−3.03	0.58
Population	−0.69	−0.11	−0.58	−0.98	−0.47	−0.51
Trend LFPR	0.11	−0.06	0.17	0.11	−0.03	0.14
Trend hours	−0.46	−0.54	0.08	−1.06	−0.83	−0.23

Notes: The table reports full sample mean revisions of estimates of trend labor and trend labor components. "Population" refers to the working-age population.

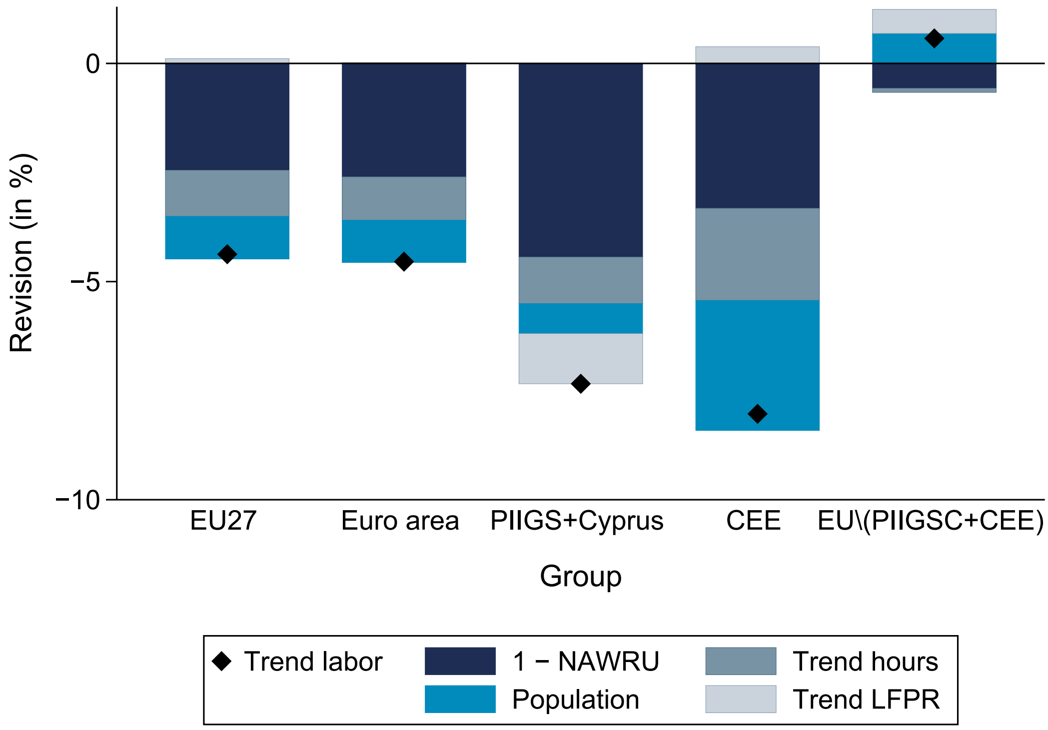

Figure 3.9. Decomposed trend labor revisions across groups for 2012

Notes: The plots show the mean contribution of component revisions to the revisions of trend labor estimates for different country groups for 2012. "Population" refers to the working-age population.

4

Improving the Reliability of Output Gap Estimates in Real Time

4.1 Introduction

The output gap measures the deviation of output from its potential and is a well-established indicator for the degree of slack in an economy. Estimates of the output gap and potential output are particularly relevant for fiscal and monetary policy. Based on the output gap, central banks decide on the interest rate level, and the evaluation of fiscal deficits with respect to potential output growth can indicate whether government spending paths are sustainable. However, neither the output gap nor potential output are observable and estimates in real time are often unreliable (e.g., Orphanides and van Norden, 2002; Orphanides, 2003). The estimation of output gaps has therefore been of major interest ever since.

The definition of the output gap often depends on the context and model (for a good overview, see, e.g., Kiley, 2013; Álvarez and Gómez-Loscos, 2018). This chapter deals with two characterizations. First, it uses output gap estimates from the Congressional Budget Office (CBO) which estimates potential output by a production function approach. Potential output therefore corresponds to the maximum sustainable production given current technologies and long-run levels of capital and labor. Second, this chapter uses an unobserved components (UC) model to decompose gross domestic product (GDP) into a cyclical and trend component by explicitly modeling macroeconomic relationships. Following the seminal paper by Okun (1962), potential output can then be characterized as a supply-side capacity measure representing the production at maximum sustainable employment without inflationary pressure.

UC models have a long tradition in the literature. Harvey (1985), Watson (1986), and Clark (1987) first suggested a univariate model which links observable output to the latent trend and cycle. Based on their original contribution, following studies added economic relationships such as Okun's law (e.g., Clark, 1989) and a Phillips curve (e.g., Kuttner, 1994; Gerlach and Smets, 1999) to name the two most prominent extensions. Current UC models often feature both relationships extended by survey-based measures like the capacity utilization rate (e.g., Benes et al., 2010; Alichi et al., 2017).

Recent contributions provide evidence that including information on the financial cycle can make output gap estimates more reliable (e.g., Borio et al., 2014, 2017; Melolinna and Tóth, 2019) and help to explain revisions in potential output (as shown in Chapter 2). Using a multivariate Hodrick-Prescott (HP) filter, Borio et al. (2014, 2017) document that potential output for the United States (US) is estimated more precisely, especially in real time. Melolinna and Tóth (2019) show that the inclusion of a financial cycle index improves real-time estimates of the output gap for the United Kingdom.

This chapter contributes to the literature by analyzing output gap estimates from various UC models which are estimated for US data by Bayesian techniques. Specifically, I outline a model featuring Okun's law, a Phillips curve, and a structural relationship between the output gap and the capacity utilization rate in manufacturing. Due to the model's flexibility, I consider several reduced model versions and show that reliable output gap estimates do not necessarily depend on model complexity but the included economic relationship. Okun's law is thereby key for reliable estimates. In addition, I document that output gap estimates only improve marginally if a model including credit growth is considered. In a real-time exercise, the inclusion of credit growth however leads to forecasts which are significantly closer to ex-post CBO estimates.

The remainder of the chapter is structured as follows. Section 4.2 outlines the baseline UC model and compares the model framework to other UC models in the literature. In Section 4.3, the data set and estimation approach are explained. Section 4.4 discusses the results for the baseline model and alternative models and analyzes the reliability of output gap forecasts in real time. Section 4.5 concludes.

4.2 A Unifying Unobserved Components Model

The baseline model is a small UC model featuring four observable variables: real GDP y_t (measured in logs times 100), the unemployment rate u_t, the inflation rate π_t, and the capacity utilization rate in manufacturing m_t. The model shares similarities to the models described by Benes et al. (2010), Alichi et al. (2017), and Melolinna and Tóth (2019), but also differs in some aspects which are explained below.

The core of the model is based on the simple assumption that every observable variable x_t, $x = \{y, u, \pi, m\}$, can be additively separated into two unobserved components: a trend $\bar{x}_t$ and cycle $\hat{x}_t$. Formally, this relationship is described by:

$$x_t = \bar{x}_t + \hat{x}_t \, . \tag{4.1}$$

The advantage of this representation is immediately clear if one considers real GDP as observable variable. Then, $\hat{y}_t$ directly corresponds to the output gap (measured in %), the main variable of interest in this chapter. To add structure to the model, the latent variables $\bar{x}_t$ and $\hat{x}_t$ are assumed to evolve according to the following state equations which describe the transition of unobserved variables from $t-1$ to t.

Following the seminal work by Harvey (1985) and Clark (1987), the trend component of GDP is assumed to follow a random walk with a stochastic drift and the output gap is described by a stationary AR(2) process. Formally, these relationships are expressed by the following three transition equations:

$$\bar{y}_t = \bar{y}_{t-1} + d^{\bar{y}}_{t-1} + \varepsilon^{\bar{y}}_t \tag{4.2}$$

$$d^{\bar{y}}_t = d^{\bar{y}}_{t-1} + \varepsilon^{d^{\bar{y}}}_t \tag{4.3}$$

$$\hat{y}_t = \phi_1 \hat{y}_{t-1} + \phi_2 \hat{y}_{t-2} + \varepsilon^{\hat{y}}_t \tag{4.4}$$

where $d^{\bar{y}}_t$ corresponds to the unobserved drift variable, ϕ_1 and ϕ_2 represent the AR(2) parameters, and $\varepsilon^{\bullet}_t$ are uncorrelated i.i.d. error terms with zero mean.

For the other three observable variables $z = \{u, \pi, m\}$, I take a parsimonious approach. The trend is assumed to evolve according to a simple I(1) process and the cycle is described by a stationary AR(1) process extended by the contemporaneous output gap. Formally, these transition equations are:

$$\bar{z}_t = \bar{z}_{t-1} + \varepsilon^{\bar{z}}_t \,, \tag{4.5}$$

$$\hat{z}_t = \left(1 - \psi^z_1\right) z_{ss} + \psi^z_1 \hat{z}_{t-1} + \psi^z_2 \hat{y}_t + \varepsilon^{\hat{z}}_t \,. \tag{4.6}$$

Again, error terms are assumed to be uncorrelated i.i.d. shocks with zero mean. The AR(1) parameter is ψ^z_1 and the parameter linking the output gap to the cycle of variable z is ψ^z_2. z_{ss} is the steady-state level of variable z and facilitates the estimation.[1] Equation (4.6) relates variable z to the output gap and thereby resembles several well-documented macroeconomic relationships. For the unemployment rate, the equation corresponds to Okun's law (Okun, 1962). For inflation, it resembles the traditional backward-looking Phillips curve (e.g., Ball and Mazumder, 2011). And for the capacity utilization rate in manufacturing, it links the output gap to cyclical fluctuations in production which are closely related to the business cycle (e.g., Morley and Piger, 2012).

There are several reasons for the specific forms of equations (4.5) and (4.6). First, the trend component has been modeled as I(2) process (e.g., Benes et al., 2010; Alichi et al.,

[1] Output gap estimates are unchanged if I use the alternative equations $\bar{z}_t = (1-\eta)\, z_{ss} + \eta \bar{z}_{t-1} + \varepsilon^{\bar{z}}_t$ and $\hat{z}_t = \psi^z_1 \hat{z}_{t-1} + \psi^z_2 \hat{y}_t + \varepsilon^{\hat{z}}_t$ with η close to one. The original equations however do not require the additional parameter η and estimation is faster.

Table 4.1—Alternative models within the model framework

ID	Variables	Related to
(A1)	y	Harvey (1985) and Clark (1987)
(A2)	y, u	Clark (1989)
(A3)	y, π	Kuttner (1994)
(A4)	y, m	–
(A5)	y, u, π	Apel and Jansson (1999)
(A6)	y, u, m	–
(A7)	y, π, m	–
(B)	y, u, π, m	Benes et al. (2010), Alichi et al. (2017)

Notes: (B) refers to the baseline model.

2017) as well as I(1) process (e.g., Melolinna and Tóth, 2019). My choice of simply using a stochastic trend is based on Bayesian model comparison. Models including a stochastic drift, i.e., featuring an I(2) process, show a smaller log-likelihood than their respective counterparts. Using the interpretation of Bayes factors from Kass and Raftery (1995), there is strong evidence against the hypothesis that the trend of any of these variables features a stochastic drift.[2] Second, equation (4.6) has been extended by the output gap lagged by one period (e.g., Melolinna and Tóth, 2019) as well as the contemporaneous output gap (e.g., Benes et al., 2010; Alichi et al., 2017). I use the latter version because it maintains the original timing in Okun's law and the traditional backward-looking Phillips curve.

The uniform model framework easily allows the estimation with only a specific subset of observable variables. This has two distinct advantages. First, the estimation on different variable sets can show which macroeconomic relationships are relevant to produce meaningful output gap estimates. Second, this approach allows the comparison of reduced models to similar models suggested in the literature.

Seven alternative models are available if GDP is a required observable variable. Table 4.1 lists the alternative models (A1)–(A7) and the baseline model (B). In addition, the reference to a prominent relative of the model is provided if such a relative exists. The table is sorted ascending by the number of included variables.

Model (A1) essentially collapses to equations (4.1)–(4.4) and thereby exactly matches the original formulation by Harvey (1985) and Clark (1987). Watson (1986) estimates a very similar model but includes a deterministic instead of stochastic drift. Model (A2) is the bivariate model proposed by Clark (1989), and model (A3) is close to the model by Kuttner (1994) with the difference that in his baseline model the cyclical component of inflation does not follow an AR(1) but MA(3) process. He shows however that results are

[2] Formally, I estimate bivariate models including observable variables y_t and z_t with and without a stochastic drift component. The log Bayes factors for the models including unemployment, inflation or capacity utilization are 3.0, 9.1, and 1.2, respectively.

very similar if an AR process is used. Model (A5) is potentially closest to Apel and Jansson (1999). The most notable difference is that the authors link the Phillips curve to the deviation of unemployment from its trend and not to the output gap directly. The capacity utilization rate is rarely included and has so far not been employed in two- and three-variable models to the best of my knowledge. It has however been used to assess output gap estimates (e.g., Camba-Mendez and Rodriguez-Palenzuela, 2003; Morley and Piger, 2012).

4.3 Estimation Strategy

In this section, I describe the macroeconomic data which enter the model as observable variables and how the model is estimated using a Bayesian approach.

4.3.1 Data

I obtain time series for real GDP, the unemployment rate, the inflation rate, and the capacity utilization rate in manufacturing for the US from the "Real-Time Data Set for Macroeconomists" provided by the Federal Reserve Bank of Philadelphia and described by Croushore and Stark (2001). The core feature of a real-time data set is that it does not only contain one value for a specific time point but a historic sequence of values for the same time point measured at different vintages. For example, real GDP at quarterly date t is published in $t + 1$ and subsequently revised in upcoming quarters $t + k$. The data reported in $t + k$ would then be one vintage in the real-time data set. In the context of this chapter, real-time data are particularly useful for forecasting the output gap in Section 4.4.3. By using different vintages in the estimation, I can exactly account for the information set which was available at the respective publication date.

Variables in the real-time data set are available at different data frequencies (i.e., time series are monthly or quarterly) and at different vintage frequencies (i.e., data is updated once per month or once per quarter). I obtain quarterly vintages of quarterly real GDP in levels, quarterly vintages of the monthly unemployment rate in percent, and monthly vintages of the monthly consumer price index (CPI) and the monthly capacity utilization rate in manufacturing in percent. All series are seasonally adjusted. For monthly time series, I average over the three months in each quarter to obtain quarterly data. In addition, I compute real GDP growth and quarter-over-quarter inflation by taking the first difference of the respective log levels.

Finally, I construct a quarterly real-time data set by using the quarterly vintages of real GDP and the unemployment rate and the first vintage of each quarter of the CPI and the capacity utilization rate in manufacturing. By doing so, I effectively adopt the timing approach of Orphanides (2003) and Edge and Rudd (2016) such that each vintage t contains quarterly time series only up to $t-1$. Therefore, any estimate for $t+h$ with $h \geq 0$ corresponds to the $(h+1)$-step ahead forecast, i.e., the forecast for $t+h$ based on information up to $t-1$.

In the following sections, subsets of vintages published between 2006Q1 and 2020Q1 will be considered. The sample period for each vintage published in t ranges between 1981Q3 and $t-1$. The first data point is carefully chosen. By using observations from 1981Q3 onwards, I exclude the high-inflationary period before Paul Volcker became chairman of the Federal Reserve in August 1979 and the subsequent short period of exceptional disinflation. This approach is similar to the one in Clarida et al. (2000) who show that monetary policy rules significantly differ across the pre-Volcker period and the Volcker-Greenspan period and likely have distinct impacts on the output gap. In addition, the exclusion is important because several studies show that the Phillips curve relationship substantially changed in the 1980s (e.g., Ball and Mazumder, 2011; Blanchard et al., 2015; Blanchard, 2016) which may lead to unstable parameter estimates in the transition equation for the cyclical component of inflation.

4.3.2 Bayesian Estimation Approach

The UC model can be expressed in state-space form with state equation (4.7) and measurement equation (4.8):

$$x_{t+1} = Fx_t + \omega_{t+1}, \quad \omega_t \overset{iid}{\sim} \mathcal{N}(0, Q) \tag{4.7}$$

$$z_t = Hx_t \tag{4.8}$$

Equation (4.7) contains the stacked equations (4.2)–(4.6) and equation (4.8) stacked versions of equation (4.1) for each observable variable. Since the model does not feature measurement error, the measurement equation does not contain any error terms.

The diffuse Kalman filter is used to evaluate the likelihood function of the model represented by equations (4.7) and (4.8). Following Koopman and Durbin (2003), the diffuse filter version is required because several transition equations feature a unit root which makes the standard initialization impossible. Parameters are estimated by Bayesian methods using MATLAB and Dynare (see Adjemian et al., 2022). Taking a Bayesian approach is beneficial because prior beliefs about the model parameters help to narrow down the large high-dimensional parameter space. Although maximum likelihood estimation is a potential alternative, An and Schorfheide (2007) point out that this traditional approach often leads to parameter estimates which are at odds with information outside the model.

Table 4.2 reports the parameter prior and posterior values. The AR(2) parameters of the output gap equation are constructed from auxiliary parameters r_1 and r_2 which are uniformly distributed on the interval $[0, 1)$. By imposing $\phi_1 = r_1 + r_2$ and $\phi_2 = -r_1 r_2$, it is

Table 4.2—Parameter prior and posterior values

	Prior			Posterior			
Parameter	Dist.	Mean	SD	Mean	SD	$\underline{\text{HPD}}$	$\overline{\text{HPD}}$
r_1	$\mathcal{U}$	0.50	0.289	0.789	0.092	0.641	0.936
r_2	$\mathcal{U}$	0.50	0.289	0.788	0.092	0.641	0.937
ψ_1^u	$\mathcal{U}$	0.50	0.289	0.704	0.039	0.641	0.769
ψ_2^u	$\mathcal{N}$	−0.50	0.200	−0.278	0.026	−0.319	−0.235
ψ_1^π	$\mathcal{U}$	0.50	0.289	0.180	0.093	0.018	0.319
ψ_2^π	$\mathcal{N}$	0.50	0.200	0.088	0.044	0.018	0.159
ψ_1^m	$\mathcal{U}$	0.50	0.289	0.160	0.080	0.024	0.283
ψ_2^m	$\mathcal{N}$	1.50	0.200	1.704	0.140	1.471	1.931
$\varepsilon^{\hat{y}}$	$\mathcal{IG}$	1.00	∞	0.368	0.034	0.312	0.422
$\varepsilon^{\bar{y}}$	$\mathcal{IG}$	0.25	∞	0.335	0.032	0.283	0.387
$\varepsilon^{d^{\bar{y}}}$	$\mathcal{IG}$	0.10	∞	0.111	0.022	0.075	0.145
$\varepsilon^{\hat{u}}$	$\mathcal{IG}$	0.10	∞	0.068	0.024	0.031	0.106
$\varepsilon^{\bar{u}}$	$\mathcal{IG}$	0.10	∞	0.097	0.027	0.049	0.136
$\varepsilon^{\hat{\pi}}$	$\mathcal{IG}$	0.50	∞	0.436	0.031	0.386	0.486
$\varepsilon^{\bar{\pi}}$	$\mathcal{IG}$	0.10	∞	0.081	0.032	0.033	0.128
$\varepsilon^{\hat{m}}$	$\mathcal{IG}$	0.20	∞	0.113	0.042	0.051	0.175
$\varepsilon^{\bar{m}}$	$\mathcal{IG}$	0.10	∞	0.544	0.053	0.458	0.633

Notes: "Dist." refers to the distribution which can be uniform ($\mathcal{U}$), normal ($\mathcal{N}$) or inverse gamma ($\mathcal{IG}$). "SD" reports the standard deviation. "$\underline{\text{HPD}}$" and "$\overline{\text{HPD}}$" refer to the lower and upper bound of the 90% highest probability density (HPD) interval, respectively.

effectively ensured that the AR(2) process is stationary.[3] The posterior means of $\phi_1 = 1.58$ and $\phi_2 = -0.62$ are in a reasonable range. Other AR(1) parameters are assumed to be uniformly distributed on the unit interval, too. These priors are essentially uniformative such that the parameter estimate is entirely driven by the data. In addition, this approach makes the estimation more flexible if it is rerun on different observable variables and data vintages. The AR(1) process for unemployment is fairly persistent whereas the processes for inflation and capacity utilization are not. For inflation, the small coefficient documents a rather flat Phillips curve and is in the range of reported estimates (e.g., Blanchard, 2016). The parameters which link the output gap to the other cycles are assumed to be normally distributed. The prior means are computed from the contemporaneous relationships between the cyclical components of HP-filtered GDP and the respective other variable. The standard deviation is chosen such that the 95% confidence interval covers values ±0.4 around the mean. The posterior parameter means can be interpreted as follows. An increase of the output gap by one percentage point reduces the cyclical component of unemployment by −0.28 percentage points and increases cyclical inflation and capacity utilization by 0.09 and 1.70 percentage points, respectively. As it is standard in the literature, shocks are assumed to

[3] An alternative approach to impose stationarity is the use of a trigonometric specification (Harvey and Jaeger, 1993). Results are very similar under this alternative specification.

be inverse gamma distributed. Prior means for standard deviations of shocks are similar to the values reported by Alichi et al. (2017). In general, shocks to the cyclical component are assumed to be larger.

Posterior distributions are constructed using the Metropolis-Hastings (MH) algorithm with two parallel chains and two million replications per chain. The first one million draws are considered as burn-in and the scale parameter of the jumping distribution's covariance matrix is tuned to obtain an acceptance ratio of 30%.

Finding the optimal posterior mode is very important because draws from the posterior distributions are only valid from that part of the Markov chain that converged. Although the Markov chain Monte Carlo (MCMC) algorithm should converge to the posterior mode at some point in time, the number of steps until this point is reached is unknown. The above estimation is therefore preceded by a three-step sequence of numerical optimizers which aims to give sufficient evidence that the optimal posterior mode is found. The posterior mode from the third step is then fed into the estimation. First, I use the evolutionary global optimizer by Hansen and Kern (2004). Second, I confirm that the posterior mode obtained from the first step is indeed an optimum by using Christopher Sims' csminwel algorithm. Third, I use Dynare's MH optimization routine to tune the scale parameter of the jumping distribution's covariance matrix.

4.4 Results

This section effectively takes the point of view of a policy maker and addresses three distinct aspects. First, I investigate how well model estimates of the output gap reflect estimates of the CBO and how estimates from the baseline model compare to estimates from reduced model versions. The full-sample results provide first indication how suitable the baseline model is to estimate the output gap. Second, I analyze whether the model extended by credit growth—which several studies suggest to incorporate—produces more reliable estimates of the output gap. Third, I examine the forecast performance of the model. Since both fiscal and monetary policy depend on sound estimates of the output gap, the model should produce robust forecasts.

4.4.1 The Baseline Model and Alternatives

The left panel of Figure 4.1 shows estimates of the output gap from the baseline four-variable model along with output gap estimates reported by the CBO for the US between 1981Q3 and 2019Q4. The CBO estimates were published in 2020Q1 such that they are directly

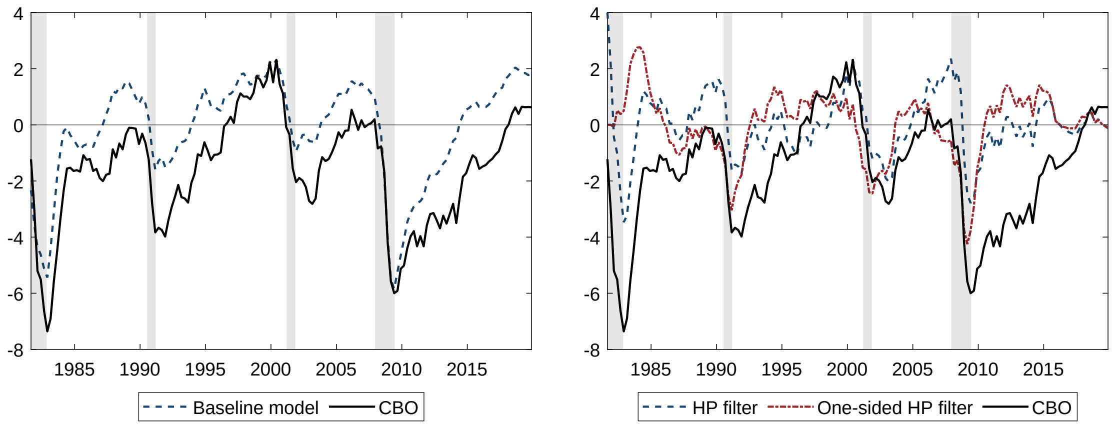

Figure 4.1. Output gap estimates from the CBO, the baseline model, and simple filters

Notes: Shaded areas report recessions identified by the National Bureau of Economic Research.

comparable.[4] The estimated output gap from the UC model matches the evolution of CBO estimates remarkably well. The correlation is very high at 0.93. Not surprisingly, model estimates tend to be larger than CBO estimates. The mean of the CBO output gap is −1.6% in the sample, whereas the model is not able to capture this large deviation from zero. The root mean squared error (RMSE) between the estimates is 1.46. The standard deviation of the CBO output gap is slightly larger which explains why estimates from the UC model are often larger (smaller) at troughs (peaks) except for the trough during the Great Recession.

By definition, the relationship between CBO estimates and the baseline model are imperfect. The CBO uses a production function approach which is highly disaggregated with respect to industry sectors (CBO, 2001, 2014). The different approach also leads to a slightly different view on the output gap. Whereas the output gap in the UC model represents the deviation of output from a sustainable trend without inflationary pressure, the CBO output gap corresponds to a deviation of output from a maximum sustainable production at current technologies and trend levels of capital and labor.

The estimation of the UC model requires a substantially larger computation time and more data input than simple filters like the HP filter and the one-sided version of the HP filter developed in Stock and Watson (1999). In common software, functions of both filters are evaluated within seconds. In addition, they require little input since only GDP and a smoothing parameter are needed to decompose GDP into a cyclical and trend component. A comparison between CBO output gaps and estimates of these filters can therefore show whether the estimation of the UC model is worth its time or researchers should use user-friendly alternatives.

---

[4] The CBO output gap corresponds to real GDP divided by potential GDP minus 1 times 100. It is computed from the two ALFRED series real GDP (GDPC1) and real potential GDP (GDPPOT) as of January 30, 2020.

The HP filter is one of the most common mathematical techniques to detrend a time series, but often exhibits spurious estimates at the sample start and end. The one-sided HP filter is advantageous in real-time applications. In contrast to the two-sided HP filter, today's value does not artificially depend on future values but only on past values in the one-sided HP filter.

The right panel of Figure 4.1 depicts CBO output gaps and the output gap estimates from the two- and one-sided HP filter using the standard smoothing parameter of $\lambda = 1600$ for quarterly data. Several differences between the estimates of the CBO and the filters are clearly visible and make estimates from the UC model superior. First, HP-filtered estimates do not closely follow the CBO estimates. Indeed, the correlations between the CBO estimates and the estimates from the two and one-sided HP filter are only 0.69 and 0.20, respectively. Second, the two filter versions are unable to capture the variation in CBO estimates. The standard deviation of the filtered cycles are only around 60% of the standard deviation of the CBO output gap. Hodrick (2020) reports a similar share between estimates from the HP filter and a simulated UC model in the spirit of Clark (1987). Finally, estimates from the one-sided HP filter are substantially smaller than estimates from the two-sided HP filter at the end of expansions. This finding is in line with Hamilton (2018) who shows that the one-sided HP filter overestimates the trend before recessions by its dependence on past values only and therefore underestimates the output gap at the same time.

It is a natural question to ask whether reduced versions of the baseline four-variable model are able to estimate the output gap equally well or even better. If a two-variable model produced similar estimates, the policy maker would be better off using the reduced model since model complexity, e.g., in terms of the number of equations and priors, and computation time reduces. Systematically reducing the baseline model is also advantageous in another aspect. The policy maker is thereby able to evaluate how the Phillips curve, Okun's law, and the link between capacity utilization and the output gap contribute to reliable output gap estimates.

The model framework allows for seven alternative models if GDP is a required observable variable in each alternative (see Table 4.1). For each alternative model, the estimation steps from Section 4.3.2 are repeated. The MCMC converges to its ergodic distribution in all models based on multivariate convergence diagnostics (Brooks and Gelman, 1998). Priors are unchanged.[5] Parameter posteriors are very similar to the baseline model across alternative models (see Table 4.6 in the Appendix.) Only model (A3), which includes GDP and inflation, shows substantial smaller AR(2) parameters for the GDP cycle.

---

[5] For models (A1), (A5), and (A6), the optimization routine by Hansen and Kern (2004) is unable to find a global optimum. Posterior modes are therefore obtained directly from Dynare's MH optimization routine. Since the MCMC converges, a stable optimum is reached.

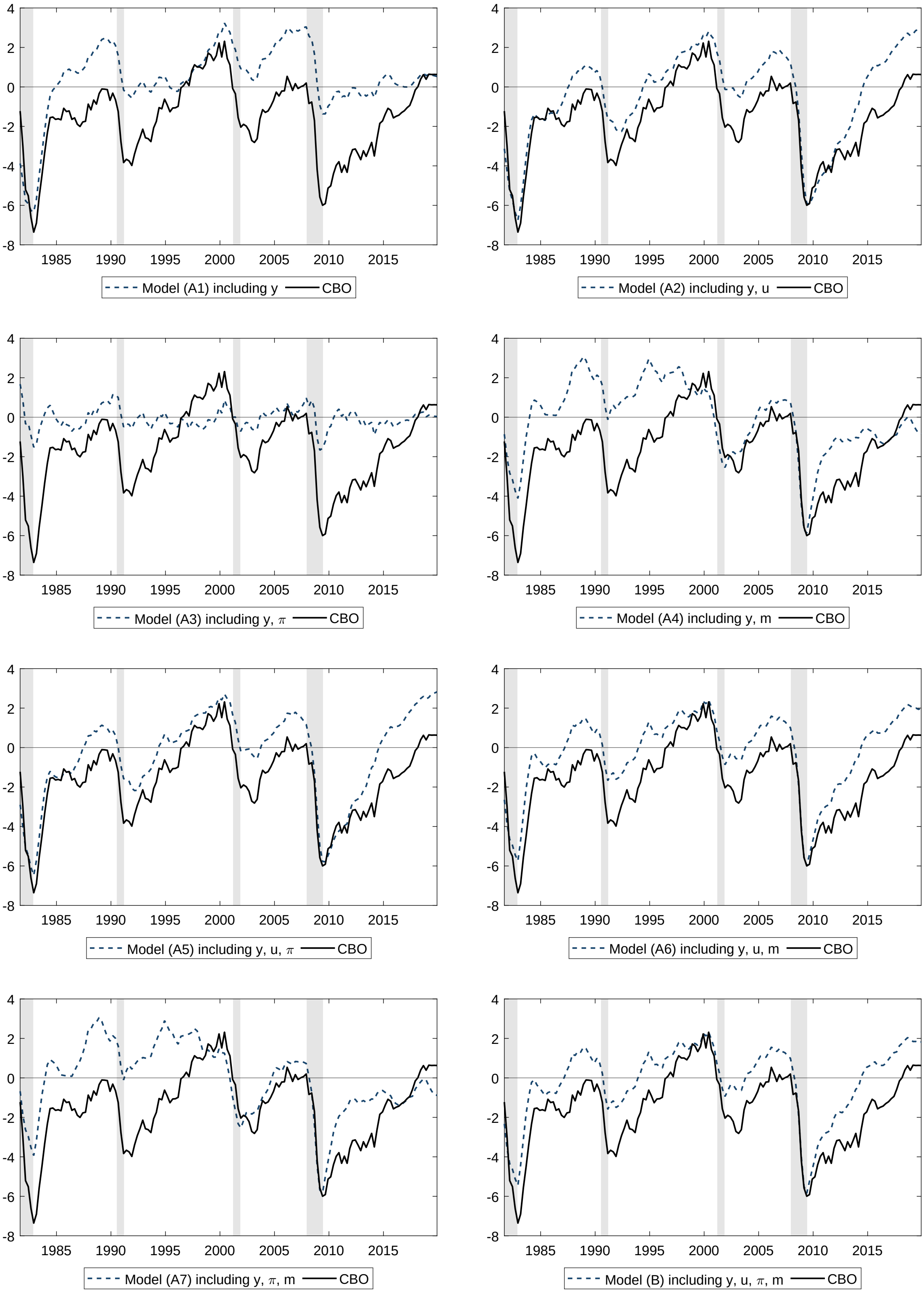

Figure 4.2. Output gap estimates from alternative models compared to the CBO

Notes: Shaded areas report recessions identified by the National Bureau of Economic Research.

Table 4.3—Comparison of output gap estimates across models

Model		Ex-post estimates				Forecasts	
		Mean	SD	Corr.	RMSE	Corr.	RMSE
(A5)	y, u, π	−0.37	2.24	0.934	1.458	0.949	1.269
(A2)	y, u	−0.39	2.34	0.932	1.477	0.947	1.367
(A6)	y, u, m	−0.19	1.95	0.930	1.564	0.907	0.636
(B)	y, u, π, m	−0.16	1.85	0.929	1.588	0.912	0.567
(HP2)	HP filter	0.00	1.15	0.690	2.120	0.995	0.387
(A4)	y, m	−0.07	1.84	0.682	2.134	−0.956	1.006
(A7)	y, π, m	−0.07	1.81	0.670	2.141	−0.950	1.022
(A3)	y, π	−0.07	0.52	0.512	2.295	0.036	0.530
(A1)	y	0.44	1.81	0.763	2.399	0.724	0.358
(HP1)	One-sided HP filter	0.00	1.20	0.203	2.598	0.383	0.798
CBO		−1.58	1.94				

Notes: Measures for estimates of the ex-post output gap are based on 154 observations in the time period 1981Q3–2019Q4. "SD" reports the standard deviation, and "Corr." refers to the correlation between the output gap estimate from the specific model and the estimate of the CBO. The RMSE compares the estimate from the model with the CBO estimate. Measures for the forecasts are based on 20 observations in the time period 2020Q1–2024Q4. Forecasts for the two- and one-sided HP filter are computed by including CBO forecasts of GDP.

Figure 4.2 depicts output gap estimates from the alternative models. Visually, output gap estimates from models which include Okun's law are much closer to the CBO estimates than others. Lower RMSE values provide evidence for this finding. Table 4.3 reports the statistics for mean estimates of the baseline model, the alternative models, and the two versions of the HP filter.[6] Models are ordered ascending with respect to the RMSE of ex-post output gap estimates, i.e., estimates for the output gap over the full sample 1981Q3–2019Q4. The RMSE for ex-post output gap estimates ranges from 1.46 for the three-variable model including GDP, unemployment and inflation to 2.60 for the one-sided HP filter. Correlation is for all models between 0.93 and 0.20. The order directly shows that output gap estimates do not necessarily improve with model complexity. For example, two- and three-variable models which solely include the Phillips curve or equations for the capacity utilization rate or both perform worse than the two-sided HP filter. Inclusion of specific variables, however, does matter. Focusing on the first four models in Table 4.3 shows that Okun's law is an important determinant for estimates of the output gap and that corresponding models without the unemployment rate have substantially lower correlation and higher RMSE. This suggests that inflation and capacity utilization are less informative for the cycle than the unemployment rate. For the models including inflation, the flattening of the Phillips curve (as documented by, e.g., Ball and Mazumder, 2011; Blanchard, 2016) can partly explain this missing relationship.

[6] Table 4.7 in the Appendix shows that results are very similar if median estimates are considered.

Blanchard (2016), for example, points out that the link between inflation and the business cycle substantially weakened between 1980 and 1990 and has remained low since then.

Table 4.3 also reports the correlation and RMSE for forecasts of the output gap.[7] While I will investigate how forecasts compare to ex-post output gap estimates in Section 4.4.3, these values provide first evidence on how close forecasts of the model are to CBO forecasts for the full sample. Correlation and RMSE are very heterogeneous across models. As expected, forecasts from the HP filter have favorable characteristics. The reason for this is, however, not a feature of the HP filter itself but is rather driven by the explanatory power of the included GDP forecasts.[8] Furthermore, the columns show that is important to evaluate both correlation and RMSE since the RMSE value could be favorable while the series is hardly or even negatively correlated. Given that a policy maker prefers a high correlation and low RMSE, the baseline model performs very well compared to its competitors.

4.4.2 Extending the Model by Credit Growth

Several empirical studies document that financial variables like credit growth comove with the business cycle in many countries (e.g., Claessens et al., 2012; Schularick and Taylor, 2012). Borio et al. (2014, 2017) therefore suggest the computation of a finance-neutral output gap to link the business to the financial cycle. Specifically, the authors incorporate credit and housing prices in a multivariate augmented version of the HP filter and show that output gap estimates for the US are estimated more precisely using this approach. Melolinna and Tóth (2019) document that output gap estimates improve for the United Kingdom if they include a financial conditions index in a three-variable UC model which is comparable to model (A5). To directly link the output gap to credit growth, I augment equation (4.4) as follows:

$$\hat{y}_t = \phi_1 \hat{y}_{t-1} + \phi_2 \hat{y}_{t-2} + \gamma \hat{w}_{t-1} + \varepsilon_t^{\hat{y}} \tag{4.9}$$

where $\hat{w}$ corresponds to the private credit growth rate at quarterly frequency.[9] Formally, the extension is similar to Melolinna and Tóth (2019) with the difference that their model features an additional measurement error in the financial variable. Estimation results are, however, very similar across the two specifications. The direct use of credit growth is advantageous because the model is thereby closer to the original findings by Claessens et al. (2012) and

[7] Forecasts for the two versions of the HP filter are computed from the ex-post GDP series extended by CBO forecasts of the growth rate of GDP.

[8] Applying the HP filter on a simple extrapolation of GDP based on the average growth rate over the last ten years yields an RMSE of 0.56 and a correlation of −0.84.

[9] Level values for credit to the private non-financial sector are obtained from the total credit statistics database of the Bank for International Settlements and correspond to the series “Q:US:P:A:M:XDC:A”. Growth is calculated by the log difference times 100 and is demeaned.

Table 4.4—Model extensions

Model		Posterior γ			Ex-post RMSE			Forecast RMSE		
		Mean	$\underline{\text{HPD}}$	$\overline{\text{HPD}}$	Orig.	Ext.	Δ	Orig.	Ext.	Δ
(A5)	y, u, π	−0.02	−0.08	0.04	1.46	1.38	0.08	1.27	1.39	−0.12
(A2)	y, u	−0.01	−0.07	0.04	1.48	1.40	0.08	1.37	1.44	−0.07
(A6)	y, u, m	−0.01	−0.07	0.04	1.56	1.51	0.05	0.64	0.77	−0.13
(B)	y, u, π, m	−0.01	−0.07	0.05	1.59	1.55	0.04	0.57	0.69	−0.12
(A4)	y, m	−0.01	−0.07	0.05	2.13	2.11	0.02	1.01	0.92	0.09
(A7)	y, π, m	−0.01	−0.06	0.05	2.14	2.12	0.02	1.02	1.03	−0.01
(A3)	y, π	0.06	−0.05	0.17	2.30	2.29	0.01	0.53	0.61	−0.08
(A1)	y	0.04	−0.09	0.17	2.40	1.86	0.54	0.36	0.47	−0.11

Notes: "$\underline{\text{HPD}}$" and "$\overline{\text{HPD}}$" refer to the lower and upper bound of the 90% high posterior density (HPD) interval, respectively. The ex-post (forecast) RMSE is based on 154 (20) observations in the time period 1981Q3–2019Q4 (2020Q1–2024Q4). "Orig." ("Ext.") refers to the RMSE for the model without (with) credit. "Δ" reports the difference between the two RMSE values.

Schularick and Taylor (2012). In addition, parameter γ will provide a direct measure on how closely private credit growth and the output gap are related.

For the prior distribution, I assume that parameter γ is normally distributed with mean and standard deviation equal to 0.25 and 1, respectively. The mean is chosen such that it matches the relationship between the cyclical component of GDP and the cyclical component of credit lagged by one period—both obtained from the HP filter. A positive mean contradicts the reported negative correlation in the literature. Therefore, the standard deviation is assumed to be large to reflect the high uncertainty of this choice.

Table 4.4 reports the posterior estimate of the coefficient γ along with the 90% highest probability density (HPD) interval, the RMSEs from the original and extended model, and the difference in RMSEs across the two models. In addition to the baseline model, I also consider the alternative models. The posterior means and HPD intervals are, in general, very similar across models suggesting a stable relationship between the output gap and credit growth. In line with the literature, the posterior mean is negative for most models but remains very small. Although the posterior mean is positive for two models, the HPD intervals are substantially larger suggesting that uncertainty around these estimates is higher. For none of the models, the posterior mean is significantly different from zero.

The results on the model fit are twofold. One the one hand, inclusion of the credit cycle improves ex-post output gap estimates numerically which is in line with the findings of Borio et al. (2014, 2017) and Melolinna and Tóth (2019). The improvement in RMSE is very similar across the models, only model (A1) benefits much more from the additional observable variable. On the other hand, however, including the credit cycle increases the RMSE between model and CBO forecasts for the majority of models and this sample. An inspection of the underlying forecasts from models including Okun's law shows that the extended versions, on average, produce more optimistic forecasts. Since the corresponding versions without credit growth also overestimate output gaps, the RMSE for the extended versions is larger. In summary, credit growth only adds little information to the output gap development for the specific sample. A potential explanation for this weak relationship could be that credit cycles are very distinct in frequency and amplitude (as, e.g., documented by Aikman et al., 2015). At this point, however, it is too early to dismiss the potential of adding credit growth to the model as such a model could still produce more reliable forecasts in real time.

4.4.3 Real-Time Forecast Performance

Applying the Kalman filter on the state space representation of the model offers a straightforward way to compute h-step ahead filtered values and forecasts (Hamilton, 1994, Ch. 13). While h-step ahead filtered values can be considered as quasi-real-time forecasts, the important difference is that they do not account for revisions in the observable variables. Croushore and Stark (2001) point out that revisions for some variables can be substantial. While unemployment revisions are negligible, real GDP, inflation, and the capacity utilization rate are revised also for several years into the past. Reasons for these revisions are for example measurement error and changes in methodology.[10] It is therefore important to take data revisions into account.

To obtain real-time forecasts, the baseline model is re-estimated on data vintages published in Q1 and Q3 between 2006 and 2019. Together with the final vintage 2020Q1, the model is therefore estimated on 29 vintages in total. In addition, the baseline model extended by credit (E), models which include Okun's law (A5, A2, A6), and the Harvey-Clark model (A1) are estimated for the same vintages.[11] Unfortunately, the Bank for International Settlements does not offer vintages of credit data. Therefore, final-vintage data is used for credit growth. Output gap forecasts of the model are compared to CBO estimates published after one year. Hence, the model forecast for the output gap in t formed h quarters before is compared to the CBO ex-post estimate for t from vintage $t+4$. The evaluation against a recent ex-post estimate is preferable because output gap estimates are often revised for many years backwards.

The latter estimate requires vintage data. I obtain real-time estimates of the output gap from CBO's "10-Year Economic Projections". Vintages are usually published twice a year in January/February (Q1) and July/August (Q3) since August 2011.[12] As a robustness exercise, model forecasts for the output gap are also compared to Greenbook ex-post output gap

[10] Tom Stark documents in the database documentation for inflation that the U.S. Bureau of Labor Statistics annually revises inflation backwards for the last five years to account for changes in seasonal factors.

[11] Models (A4), (A7), and (A3) are excluded from this exercise because of the substantially larger RMSEs and lower correlation coefficients.

[12] There were three exceptions to this rule. The CBO did not publish any projections in 2013Q3, 2017Q3, 2018Q1.

Table 4.5—Output gap forecasts compared to CBO and Greenbook estimates

Model versus		RMSE				CRPS mean			
		Total	$h = 2$	$h = 4$	$h = 8$	Total	$h = 2$	$h = 4$	$h = 8$
CBO estimates									
(A5)	y, u, π	3.19	3.88	3.73	3.50	2.03	2.73	2.22	1.75
(A2)	y, u	3.18	3.92	3.74	3.47	2.01	2.75	2.21	1.70
(A6)	y, u, m	3.16	3.77	3.62	3.45	2.03*	2.68	2.19	1.78
(B)	y, u, π, m	3.15	3.74	3.58	3.44	2.01	2.64	2.14	1.77
(E)	y, u, π, m, cr	3.13***	3.73	3.58	3.42	1.97***	2.60***	2.11	1.75
(A1)	y	4.05***	4.17	4.25	4.29	2.33	2.50	2.41	2.31
Greenbook estimates									
(A5)	y, u, π	2.86	2.20	2.71	3.65	1.77	1.60	1.64	2.13
(A2)	y, u	2.84	2.17	2.69	3.65	1.73	1.57	1.61	2.09
(A6)	y, u, m	2.91	2.27	2.76	3.67	1.88	1.78	1.75	2.18
(B)	y, u, π, m	2.92	2.29	2.77	3.68	1.88	1.79	1.75	2.18
(E)	y, u, π, m, cr	2.88***	2.25***	2.74	3.63	1.83***	1.73***	1.69***	2.13**
(A1)	y	4.41**	4.26	4.38	4.68	2.62***	2.58	2.61	2.76

Notes: "Total" reports the respective measure computed on horizons 1–8 and is based on 152 (144) observations for CBO (Greenbook) data. The number of observations for separate horizons is 20 per horizon for CBO data and 19, 18, 16 per horizon 2, 4, 8 for Greenbook data. ***, **, and * correspond to significance levels of 1%, 5%, and 10% and are obtained from a two-sided Diebold-Mariano test.

estimates for t from vintage $t + 4$. Greenbook vintages are published quarterly but are only available with a publication lag of five years, i.e., until 2016Q4.

Table 4.5 reports the total RMSE for forecasts up to eight-quarters ahead and the RMSE for horizons two, four, and eight quarters, separately. For comparison, the mean of the continuous ranked probability score (CRPS) is reported for the same horizons. The main advantage of the CRPS is that it does not only indicate how close the point forecast is to the ex-post estimate but also provides a score on how the ex-post estimate is embedded in the probability distribution of the forecast (e.g., Gneiting and Raftery, 2007).[13] Stars indicate whether the squared errors or the CRPS values are significantly different from the baseline model using the well-established test of Diebold and Mariano (1995).

Two results stand out. First, models including Okun's law produce more accurate estimates than the classical Harvey-Clark model. For both sets, the total RMSE for estimates from the Harvey-Clark model is significantly larger than the total RMSE for estimates from the baseline model. Second, the baseline model extended by credit significantly improves output gap forecasts. For Greenbook data, the CRPS is significantly smaller across all horizons. For CBO data, the total CRPS and the CRPS for the two-quarters ahead forecast are significantly smaller. This is not driven by the choice of using the ex-post estimate for t from vintage $t + 4$. In fact, results for ex-post estimates for t from vintage $t + 2$ are very similar (see Table 4.8 in

[13] Since the estimation does not offer closed-form solutions for the forecast distributions, I follow the strategy of Jordan et al. (2019) and compute the CRPS from simulated samples of 100,000 draws from the empirical distribution functions for each forecast and each horizon.

the Appendix). Although the forecast improvement is numerically small, these results shows that incorporating the credit cycle can significantly improve output gap forecasts and that credit growth contains important information about the business cycle.

4.5 Conclusion

The output gap is unobservable and estimates can vary considerably. In this chapter, I analyze how ouput gap estimates depend on structural relationships. Specifically, I outline a small four-variable UC model with three economic relationships: Okun's law, a Phillips curve, and a structural relationship between the output gap and the capacity utilization rate. Considering several reduced versions of this model, I document that Okun's law is an essential requirement to align output gap estimates with estimates from the CBO. In contrast, a Phillips curve relationship seems not helpful to produce reliable estimates of the output gap. Extending the model by credit growth—as, for example, suggested by Borio et al. (2014, 2017)—improves output gap estimates for the US only marginally. When it however comes to forecasting, a model including credit growth produces significantly better forecasts of the output gap.

4.A Appendix

Table 4.6—Parameter posterior means and standard deviations for alternative models

	(B)	(A1)	(A2)	(A3)	(A4)	(A5)	(A6)	(A7)
r_1	0.789	0.798	0.824	0.551	0.795	0.818	0.798	0.791
	(0.092)	(0.163)	(0.091)	(0.218)	(0.084)	(0.094)	(0.087)	(0.086)
r_2	0.788	0.798	0.824	0.550	0.795	0.817	0.798	0.792
	(0.092)	(0.164)	(0.090)	(0.218)	(0.085)	(0.094)	(0.087)	(0.085)
ψ_1^u	0.704		0.547			0.558	0.695	
	(0.039)		(0.068)			(0.067)	(0.039)	
ψ_2^u	−0.278		−0.338			−0.342	−0.277	
	(0.026)		(0.050)			(0.049)	(0.026)	
ψ_1^π	0.180			0.134		0.199		0.163
	(0.093)			(0.080)		(0.094)		(0.088)
ψ_2^π	0.088			0.306		0.061		0.111
	(0.044)			(0.126)		(0.037)		(0.044)
ψ_1^m	0.160				0.143		0.146	1.880
	(0.080)				(0.075)		(0.077)	(0.147)
ψ_2^m	1.704				1.861		1.705	0.153
	(0.140)				(0.149)		(0.141)	(0.077)
$\varepsilon^{\hat{y}}$	0.368	0.362	0.349	0.372	0.355	0.348	0.358	0.353
	(0.034)	(0.077)	(0.046)	(0.068)	(0.033)	(0.046)	(0.034)	(0.033)
$\varepsilon^{\bar{y}}$	0.335	0.318	0.337	0.229	0.347	0.336	0.336	0.347
	(0.032)	(0.086)	(0.038)	(0.094)	(0.033)	(0.039)	(0.031)	(0.033)
$\varepsilon^{d^{\bar{y}}}$	0.111	0.118	0.102	0.196	0.131	0.103	0.108	0.132
	(0.022)	(0.046)	(0.021)	(0.039)	(0.025)	(0.021)	(0.022)	(0.025)
$\varepsilon^{\hat{u}}$	0.068		0.077			0.078	0.071	
	(0.024)		(0.020)			(0.020)	(0.024)	
$\varepsilon^{\bar{u}}$	0.097		0.074			0.073	0.093	
	(0.027)		(0.028)			(0.028)	(0.028)	
$\varepsilon^{\hat{\pi}}$	0.436			0.421		0.443		0.432
	(0.031)			(0.029)		(0.030)		(0.029)
$\varepsilon^{\bar{\pi}}$	0.081			0.054		0.074		0.079
	(0.032)			(0.019)		(0.030)		(0.029)
$\varepsilon^{\hat{m}}$	0.113				0.286		0.111	0.292
	(0.042)				(0.080)		(0.041)	(0.074)
$\varepsilon^{\bar{m}}$	0.544				0.162		0.554	0.146
	(0.053)				(0.154)		(0.052)	(0.142)

Notes: Values in parentheses report the standard deviation around the posterior mean of the parameter. Parameter estimates in column (B) repeat the posterior mean and standard deviation for the baseline model from Table 4.2.

Table 4.7—Comparison of median output gap estimates across models

Model		Ex-post estimates				Forecasts	
		Mean	SD	Corr.	RMSE	Corr.	RMSE
(A5)	y, u, π	−0.33	2.24	0.934	1.486	0.937	1.231
(A2)	y, u	−0.36	2.34	0.932	1.502	0.939	1.291
(A6)	y, u, m	−0.18	1.94	0.930	1.569	0.904	0.616
(B)	y, u, π, m	−0.16	1.85	0.929	1.593	0.899	0.545
(HP2)	HP filter	0.00	1.15	0.690	2.120	0.995	0.387
(A4)	y, m	−0.07	1.87	0.665	2.167	−0.888	1.003
(A7)	y, π, m	−0.07	1.84	0.659	2.169	−0.923	1.027
(A3)	y, π	−0.06	0.52	0.505	2.306	0.051	0.528
(A1)	y	0.11	1.44	0.742	2.128	0.362	0.434
(HP1)	One-sided HP filter	0.00	1.20	0.203	2.598	0.383	0.798
CBO		−1.58	1.94				

Notes: Measures for estimates of the ex-post output gap are based on 154 observations in the time period 1981Q3–2019Q4. "SD" reports the standard deviation, and "Corr." refers to the correlation between the output gap estimate from the specific model and the estimate of the CBO. The RMSE compares the estimate from the model with the estimate from the CBO. Measures for the forecasts are based on 20 observations in the time period 2020Q1–2024Q4. Forecasts for the two- and one-sided HP filter are computed by including CBO forecasts of GDP.

Table 4.8—Output gap forecasts compared to CBO and Greenbook estimates at $t + 2$

Model versus		RMSE				CRPS mean			
		Total	$h = 2$	$h = 4$	$h = 8$	Total	$h = 2$	$h = 4$	$h = 8$
CBO estimates									
(A5)	y, u, π	3.39	4.67	3.91	3.52	2.24	3.28	2.47	1.82
(A2)	y, u	3.39	4.70	3.93	3.48	2.22	3.30	2.47	1.77
(A6)	y, u, m	3.33**	4.63	3.76	3.47	2.22***	3.21	2.41	1.84*
(B)	y, u, π, m	3.30	4.65	3.73	3.42	2.18	3.18	2.38	1.81
(E)	y, u, π, m, cr	3.29**	4.63***	3.73	3.41	2.15***	3.14	2.34**	1.80
(A1)	y	3.98**	4.90	4.13	4.18	2.32	2.61	2.42	2.27
Greenbook estimates									
(A5)	y, u, π	2.84	2.19	2.69	3.65	1.74	1.62	1.62	2.09
(A2)	y, u	2.81	2.15	2.65	3.64	1.70	1.58	1.58	2.04
(A6)	y, u, m	2.90**	2.28	2.76	3.68	1.85	1.78	1.75	2.15
(B)	y, u, π, m	2.92	2.30	2.78	3.69	1.86	1.79	1.75	2.15
(E)	y, u, π, m, cr	2.88***	2.26***	2.74***	3.64**	1.80***	1.73***	1.69***	2.10**
(A1)	y	4.45***	4.32	4.43	4.73	2.60**	2.56	2.61	2.75

Notes: "Total" reports the respective measure computed on horizons 1–8 and is based on 152 (144) observations for CBO (Greenbook) data. The number of observations for separate horizons is 20 per horizon for CBO data and 19, 18, 16 per horizon 2, 4, 8 for Greenbook data. ***, **, and * correspond to significance levels of 1%, 5%, and 10% and are obtained from a two-sided Diebold-Mariano test.

5

The Effect of the European Excessive Deficit Procedure on Fiscal Spending Multipliers

Joint with David A. Vespermann

5.1 Introduction

The Stability and Growth Pact is one of the core elements of European fiscal governance. Implemented in 1997, it provides a rule-based framework for fiscal policy to ensure sound public finances and fiscal discipline in the European Union (EU). The Excessive Deficit Procedure (EDP) is the corrective arm of this framework. If a member state runs an excessive deficit or accumulates high levels of public debt, an EDP is launched. The adjustments prescribed by the EDP provisions aim to ensure appropriate fiscal policies and, as a consequence, sustainable public finances. These measures put an additional strain on public budgets and fiscal spending behavior. This may interfere with the impact of government spending on economic activity and, in particular, the effectiveness of fiscal stimulus in times of weak fiscal position or recessionary episodes. In this chapter, we explore whether the EDP affects fiscal spending multipliers. Specifically, we estimate state-dependent impulse response functions and cumulative multipliers for a panel of European countries using local projections to identify how the EDP alters the transmission of government spending shocks.

In the literature on the determinants of fiscal multipliers, the recent debate has revolved around potential state dependence. In particular, a growing number of studies explores how economic conditions affect the output response to fiscal stimulus. Regarding the impact of the business cycle position, the empirical evidence is mixed. Auerbach and Gorodnichenko (2012, 2013) document that fiscal multipliers in recessions are larger than in expansions. By contrast, Owyang et al. (2013) and Ramey and Zubairy (2018) do not find evidence that public spending is more effective during times of slack measured by high levels of the unemployment rate.

Ilzetzki et al. (2013) evaluate how fiscal multipliers vary across different economic environments and find larger multipliers under fixed exchange rate regimes and negative multipliers in high-debt countries. The latter finding is confirmed by Nickel and Tudyka (2014), whereas, more recently, Banerjee and Zampolli (2019) report small but positive fiscal multipliers below one. Corsetti et al. (2012b) document negative government spending multipliers for countries with weak fiscal positions as measured by high debt levels and excessive deficits. This finding is driven by significant crowding-out of private investment. Huidrom et al. (2020) identify two channels through which high public debt has an effect on the size of fiscal multipliers. First, Ricardian households decrease consumption in response to fiscal stimulus because they anticipate adjustments in the future. Second, public spending raises sovereign credit risk and puts upward pressure on interest rates, which reduces private demand. Thus, lower multipliers in weak fiscal positions are rationalized by crowding-out of private consumption.

Surpisingly, the literature analyzing the effectiveness of the EDP is very limited. Hagen and Eichengreen (1996) suspect that the EDP is at best redundant but more likely detrimental as it limits policy makers' scope for reactions. De Jong and Gilbert (2020) show that countries in the EDP largely comply in terms of fiscal reactions with the recommendations of the European Commission. Górnicka et al. (2020) calculate implied fiscal multipliers in the EDP from these recommendations and document implied multipliers below one, which increase over time. So far, however, none has systematically estimated fiscal multipliers for countries in the EDP over several horizons, compared them to multipliers for countries not in the EDP or investigated the transmission mechanisms of the EDP on the economy. We aim to fill this gap.

We contribute to the literature by evaluating the specific impact of the EDP on fiscal spending multipliers. More precisely, we screen official documents provided by the European Commission to construct a dummy variable indicating whether a country is in the EDP or not. We use this dummy as state variable in the estimation of cumulative spending multipliers and the corresponding impulse response functions. We document that cumulative multipliers for countries in the EDP are significantly larger, suggesting that government spending in these countries is more effective. The inspection of the underlying transmission mechanisms reveals that this finding is mainly driven by a decrease in interest rates and substantial crowding-in of private investment in response to fiscal stimulus. At the same time, we observe a decrease in debt. We find that the EDP is especially effective in bad times, as indicated by larger multipliers for countries with weak fiscal positions, in recessions or during banking crises. In addition, the comparison with alternative state variables provides evidence that the EDP is not simply a proxy for such bad times. Finally, we show that fiscal multipliers are significantly understated if estimated in real time.

These results have some important policy implications. First, and most importantly, our findings are evidence that the EDP is functional. The output response to government spending is stronger and public debt decreases in response to a positive spending shock for countries in the procedure. Thus, fiscal stimulus could help these countries to return on a path towards more sustainable public finances. By contrast, fiscal consolidation would have substantial contractionary and therefore harmful effects in these economies. Second, the EDP is especially effective in bad times which is the purpose it is designed for. Third, lower-than-expected fiscal multipliers in real time mask the effect of fiscal stimulus for countries in the EDP. This is especially important for policy makers who should treat real-time multipliers cautiously.

The remainder of the chapter is structured as follows. In Section 5.2, we describe the EDP and explain the construction of the EDP dummy variable. Section 5.3 introduces the empirical strategy and the data. We discuss the estimation results in Section 5.4 and compare these results with fiscal multipliers based on alternative state variables in Section 5.5. In Section 5.6, we present results for real-time fiscal multipliers. Section 5.7 concludes.

5.2 The Excessive Deficit Procedure

The EDP is a multi-step procedure with the objective to correct imbalances of public finances by urging the affected member states to reduce excessive deficits and/or debt levels. Despite the name of the procedure, the EDP always takes into account both criteria. The details of the EDP provisions are laid down in Article 126 of the *Treaty on the Functioning of the European Union* (European Union, 2008). An EDP can be launched in two cases. First, for countries with headline budget deficits exceeding, or being at risk to exceed, the limit of 3% of GDP. Second, for countries with public debt levels above 60% of GDP which do not decrease at a sufficient pace, defined as 1/20th of the gap between the actual level and the 60% reference value per year. The initial step is a report by the European Commission notifying the non-compliance of a member state with the requirements for deficit and/or debt. In the next step, the Commission proposes whether or not an excessive deficit should be declared, taking into account "all other relevant factors, including the medium-term economic and budgetary position" (European Union, 2008) of the country. Based on this proposal, the European Council officially decides on the existence of an excessive deficit. If the Council declares that an excessive deficit exists, this decision formally opens the EDP and triggers a series of actions. Commission and Council issue recommendations on how to correct the paths of aggregate deficit and debt and set a deadline. The measures undertaken by the member states are continuously monitored and, in case of non-effective action, the Council can impose sanctions. The first stage is a compulsory non-interest-bearing deposit of 0.2% of GDP, which can be converted into a fine in the second stage. In practice, sanctions have never been imposed so far. The EDP is closed when the Council, on a recommendation by the Commission, decides that the excessive deficit has been corrected.

Table 5.1—EDPs included in the sample

#	Country	1st neg. report	Comm. rec.	Start	End
1	Belgium	–	11 Nov 2009	2 Dec 2009	20 Jun 2014
2	Czech Republic	–	24 Jun 2004	5 Jul 2004	3 Jun 2008
3	Czech Republic	7 Oct 2009	11 Nov 2009	2 Dec 2009	20 Jun 2014
4	Denmark	12 May 2010	15 Jun 2010	13 Jul 2010	20 Jun 2014
5	Finland	12 May 2010	15 Jun 2010	13 Jul 2010	12 Jul 2011
6	France	2 Apr 2003	7 May 2003	3 Jun 2003	30 Jan 2007
7	France	18 Feb 2009	24 Mar 2009	27 Apr 2009	22 Jun 2018
8	Germany	19 Nov 2002	8 Jan 2003	21 Jan 2003	5 Jun 2007
9	Germany	7 Oct 2009	11 Nov 2009	2 Dec 2009	22 Jun 2012
10	Hungary	12 May 2004	24 Jun 2004	5 Jul 2004	21 Jun 2013
11	Ireland	18 Feb 2009	24 Mar 2009	27 Apr 2009	17 Jun 2016
12	Italy	7 Oct 2009	11 Nov 2009	2 Dec 2009	21 Jun 2013
13	Netherlands	28 Apr 2004	19 May 2004	2 Jun 2004	7 Jun 2005
14	Netherlands	7 Oct 2009	11 Nov 2009	2 Dec 2009	20 Jun 2014
15	Portugal	22 Jun 2005	20 Jul 2005	20 Sep 2005	3 Jun 2008
16	Portugal	7 Oct 2009	11 Nov 2009	2 Dec 2009	16 Jun 2017
17	Slovak Republic	7 Oct 2009	11 Nov 2009	2 Dec 2009	20 Jun 2014
18	Spain	18 Feb 2009	24 Mar 2009	27 Apr 2009	14 Jun 2019
19	United Kingdom	21 Sep 2005	11 Jan 2006	24 Jan 2006	9 Oct 2007
20	United Kingdom	11 Jun 2008	2 Jul 2008	8 Jul 2008	5 Dec 2017

Notes: "1st neg. report" refers to the date at which the European Commission filed the first negative report on the non-compliance of the country. "Comm. rec." refers to the date the Commission recommends to the European Council that the EDP should be declared. "Start" and "End" denote the date of the opening and closing of the EDP as decided by the Council. For EDPs no. 1 and 2, the Commission proposed the EDP without a first negative report.

For our analysis, we use the official documents on all ongoing and closed EDPs provided by the European Commission. Out of the 27 EU countries and the former member United Kingdom, 25 countries have been at least once in the procedure between 2000 and 2019. Estonia, Luxembourg and Sweden are the only countries which have never been in an EDP. Table 5.1 lists the EDPs that are included in our sample.[1] Twelve of these EDPs were opened during the Great Recession or in its aftermath. At the end of 2019, there was no ongoing EDP. The last procedure was closed in summer 2019 after ten years of monitoring Spain. This is also the longest EDP in our sample where the average EDP duration is five years. The table shows that Commission and Council act rather swiftly with an average duration of two months between the first negative report of the Commission, the following Commission's recommendation on the opening of the EDP and the decision of the Council on that matter. After the Commission recommends an opening, the Council opens the EDP usually within one month. There exists no recommendation which was not followed by an opening of an EDP.

[1] We cannot include all EDPs which have been opened by the Council in the period 2000–2019 because of missing data for some European countries. The main obstacle is insufficient data for the construction of government spending shocks which we detail in Section 5.3.3.

We code a dummy variable that captures the periods in which a country was in an EDP. Since we use semi-annual data from OECD Economic Outlook (EO) editions, we have to make sure that changes in the dummy variable correspond to the EO editions. More specifically, the EDP dummy variable is set to 1 from period t onwards if the procedure was opened between the forecast cut-off date of EO edition $t-1$ and the forecast cut-off date of edition t. Accordingly, the dummy variable is set to 0 from period $t+1$ onwards if the procedure was closed between the forecast cut-off date of EO edition t and the forecast cut-off date of edition $t+1$. By doing so, we carefully account for the information set which is available at the time each EO edition is published. This approach provides us with country-specific EDP dummies which we use as state variables in the panel estimation of state-dependent fiscal multipliers and impulse response functions presented in the next section.

5.3 Methodology and Data

In this section, we describe our empirical strategy, our data and how we identify government spending shocks.

5.3.1 Empirical Strategy

We use the local projection method proposed by Jordà (2005) to estimate impulse response functions and fiscal multipliers directly. This method allows for a flexible specification of state dependence and does not implicitly restrict the model dynamics. In particular, we follow Auerbach and Gorodnichenko (2013) and estimate state-dependent local projections using panel data for European countries. The state-dependent responses $\beta_{0,h}$ and $\beta_{1,h}$ of the variable of interest $Z_{i,t+h}$ to an exogenous change in government spending for each horizon $h = 0, \ldots, H$ are estimated from the following regression model:

$$\begin{aligned} Z_{i,t+h} \;=\; & \left(1 - \mathcal{I}_{i,t-1}\right)\left[\alpha_{0,i,h} + \beta_{0,h} G_{i,t} + \Phi_{0,h}(L) X_{i,t-1}\right] \\ & + \mathcal{I}_{i,t-1}\left[\alpha_{1,i,h} + \beta_{1,h} G_{i,t} + \Phi_{1,h}(L) X_{i,t-1}\right] + \sum_{k=1}^{2} \psi_k T_t^k + \varepsilon_{i,t+h}\,, \end{aligned} \tag{5.1}$$

where $\mathcal{I}_{i,t-1}$ indicates the state of the economy of country i in the period before the change in government spending, $\alpha_{\bullet,i,h}$ measures unobserved state-dependent fixed effects, $G_{i,t}$ represents government spending in the current period, L refers to the lag operator, $X_{i,t-1}$ is a vector of controls and the series of ψ_k captures a time trend. In our baseline specification, the state indicates whether a country is in an ongoing EDP. The set of control variables includes output, government spending, private consumption, private investment, the interest-rate spread, the marginal tax rate and the public debt level. The variable of interest $Z_{i,t+h}$ is a variable from this set. The units of all variables measured in levels (i.e., output, goverment spending, private consumption, private investment and the public debt level) are normalized by an estimate of trend GDP (Gordon and Krenn, 2010; Ramey and Zubairy, 2018) rather than lagged GDP (as done by, e.g., Hall, 2009; Barro and Redlick, 2011). The latter approach produces fiscal multipliers which vary over the business cycle. We obtain trend GDP from a polynomial of order 3.[2] In order to address potential endogeneity issues in our regressions, we use an instrumental variable approach: Normalized government spending $G_{i,t}$ is instrumented by the forecast error of government spending (Auerbach and Gorodnichenko, 2012, 2013). We discuss the details on the identification of government spending shocks in Section 5.3.3. Finally, we use robust standard errors that account for cross-sectional dependence and autocorrelation as proposed by Driscoll and Kraay (1998).

The impulse response function for each of the states is constructed using the sequence of responses $\{\beta_{\bullet,h}\}_{h=0}^{H}$. The impulse responses trace the impact of the exogenous shock on the path of specific variables and reveal the underlying transmission mechanisms. Following Mountford and Uhlig (2009) and Ramey and Zubairy (2018), we argue that the policy-relevant measure for the aggregate effect of government spending shocks on the economy is given by the cumulative fiscal spending multiplier. The cumulative multiplier compares the cumulative output response (i.e, the integral of the output response) to the cumulative path (i.e., the integral) of government spending, thereby providing a measure for the impact of fiscal stimulus over time. Note that this definition is different from the one in Blanchard and Perotti (2002) and Auerbach and Gorodnichenko (2012, 2013) who report fiscal multipliers given by the peak response of output relative to the initial fiscal spending impulse. Peak multipliers, however, do not take into account the underlying response of government spending and therefore complicate the comparison across estimations.

Specifically, we follow Ramey and Zubairy (2018) and Bernardini and Peersman (2018) and estimate state-dependent cumulative fiscal multipliers for horizon h in one step using the following regression:[3]

$$\begin{aligned}\sum_{j=0}^{h} Y_{i,t+j} \;=\; & \left(1-\mathcal{I}_{i,t-1}\right)\left[\alpha_{0,i,h}+M_{0,h}\sum_{j=0}^{h}G_{i,t+j}+\Phi_{0,h}(L)X_{i,t-1}\right] \\ & +\mathcal{I}_{i,t-1}\left[\alpha_{1,i,h}+M_{1,h}\sum_{j=0}^{h}G_{i,t+j}+\Phi_{1,h}(L)X_{i,t-1}\right]+\sum_{k=1}^{2}\psi_k T_t^k+\varepsilon_{i,t+h}\,,\end{aligned} \tag{5.2}$$

[2] Ramey and Zubairy (2018) use a polynomial of order 6. We justify our choice by the smaller sample period which makes less turning points necessary.

[3] Ramey and Zubairy (2018) argue that the one-step estimation procedure is identical to a three-step procedure in which the sum of the output responses is divided by the sum of government spending responses if all responses are estimated on the same sample.

where $Y_{i,t+j}$ denotes normalized output and all other variables are defined as explained above. We instrument $\sum_{j=0}^{h} G_{i,t+j}$ by the forecast error of government spending. Hence, the instrument is independent of the horizon h, see Ramey and Zubairy (2018). In this specification, the estimated coefficients $M_{0,h}$ and $M_{1,h}$ provide direct measures for the cumulative fiscal multipliers for each state at horizon h.

5.3.2 Data

The sample covers the period 2000H1–2019H1 and 17 European countries.[4] Our main data source is the OECD EO published on a biannual basis (spring and autumn of each year). We use this data set for two reasons. First, because of its large coverage. It includes macroeconomic variables along with forecasts of up to two years ahead for most European countries. Hence, this provides us with a consistent data set. Second, this source additionally provides us with government spending forecasts from each EO edition which we use to identify government spending shocks, see Section 5.3.3 for details. Since the EO reports quarterly values only starting from edition 73 (published in spring 2003) and semiannual values before, we harmonize the frequency of all variables to consistently use semiannual data in the analysis.

We mainly use data from EO edition 106 published in autumn 2019. We take real data for output, government spending (government consumption plus—if available—government investment), private consumption and private investment.[5] We construct semiannual levels for these variables by aggregating the values from Q1 and Q2 (Q3 and Q4) for H1 (H2).[6] For the interest-rate spread, we take the 10-year government bond spread vis-à-vis Germany (in percent). The value for the semester is given by the average over its two quarters. In addition, we take the public debt level which is measured at the end of the year and reported at annual frequency only. Therefore, we use the annual value of year t for H2 in t and H1 in $t + 1$.

Finally, we take the annual marginal personal income tax rate (in percent) from Table I.4 of the OECD Tax Database at an income level of 100% of the average wage. This rate includes the central government and sub-central income tax plus the employee social security contributions. We use the value of year t for H1 and H2 in t.

[4] Our sample includes Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Slovakia, Spain, Sweden and the United Kingdom.

[5] Government and private investment are only reported for seven countries separately. For the other ten countries in our sample the sum of the two is reported. We therefore include government investment where it is available for these seven countries and proxy private investment by the sum of the annual shares of household and corporate investment multiplied by semiannual total investment. The latter approach is valid. The correlation between implied private investment and reported private investment is very high (above 0.85 for each of the seven countries).

[6] Before aggregation, we replace government investment for the United Kingdom in 2005Q2 by the average of the previous quarter and the following quarter because government investment was exceptionally negative due to the transfer of nuclear reactors to the government.

5.3.3 Shock Identification

We define government spending shocks as the forecast error of government spending growth. By doing so, goverment spending shocks aim to measure the unexpected change in government spending growth and can be used as an instrument for government spending. This approach was put forward by Auerbach and Gorodnichenko (2012, 2013) and has been widely used since then. The identification is based on the timing assumption of Blanchard and Perotti (2002). That is, unexpected changes in government spending growth are exogenous and therefore not a contemporaneous response to macroeconomic aggregates.

The approach can be divided into two steps. First, we calculate government spending growth from real government consumption plus—if available—real government investment for all countries and all EO editions between edition 67 (published in spring 2000) and edition 106 (published in autumn 2019). The inclusion of government investment is important because government investment multipliers are much smaller as shown by Boehm (2020). Solely using government consumption could consequently exaggerate the fiscal multipliers. We refer to the calculated growth rates by $g^s_{i,t}$. This is the semiannual government spending growth rate for country i between semesters $t-1$ and t based on data from the EO edition published in semester s. Second, we calculate the forecast error of government spending growth by:

$$FE_{i,t} = g^{2019H2}_{i,t} - g^{t-1}_{i,t}, \tag{5.3}$$

where $g^{2019H2}_{i,t}$ is the realized growth rate from EO edition 106 and $g^{t-1}_{i,t}$ is the one-step-ahead forecast for semester t published in semester $t-1$. We make sure that the growth rates are comparable in terms of the inclusion of government investment, i.e., either both growth rates contain government investment or both do not.

There are several alternative choices for the realized growth rate as the first release of the growth rate is obviously published in $t+1$ and one can consider realizations from $t+1$ to the latest available semester.[7] The OECD continuously revises government consumption and government investment also back to the past and we do not know which realization forecasters aimed to predict. Even if one arbitrarily chose the realizations after a specific fixed horizon h (i.e., the realization of t published in $t+h$), it would be uncertain whether forecasters consistently aimed to predict the realization published in $t+h$ for all t. We do not want to make a stand on which growth rate should be considered as the final release, i.e., the aim of the forecaster's prediction. By using the growth rate from the latest available semester, we are agnostic about which growth rate represents the final realization. The advantage of

[7] For example, Auerbach and Gorodnichenko (2013) use the realization published in $t+4$.

this approach is that our sample size increases considerably compared to the approach by Auerbach and Gorodnichenko (2013), while we will show that our results are very similar.

5.4 Fiscal Multipliers in the Excessive Deficit Procedure

In this section, we report the results of our baseline estimation, discuss the underlying mechanisms and evaluate the EDP multiplier in bad times.

5.4.1 Baseline Results

Figure 5.1 shows the estimates for the cumulative fiscal multipliers from our baseline specification given by equation (5.2). The left panel displays the linear cumulative multipliers for the whole sample, i.e., the multiplier estimates for the case without state dependence. We plot a horizon of five semesters and the shaded area refers to the 90% confidence intervals calculated from Driscoll-Kraay standard errors. The point estimates for the linear multipliers are positive and smaller than one across all horizons. The fiscal multiplier is 0.2 on impact and slightly increases over time reaching 0.7 after two years ($h = 4$), which is in the range of unconditional estimates for European countries commonly reported in the literature, see Mineshima et al. (2014) for an extensive survey.

The right panel of Figure 5.1 displays the cumulative state-dependent multipliers. The state indicates whether a country has been in an ongoing EDP in the period before the change in government spending. The red line refers to episodes in which countries are in an EDP, whereas the blue line is associated with countries which are not. The red dashed lines and the shaded area again indicate the 90% confidence intervals. The state-dependent multipliers for the EDP sample are positive and larger than one across all horizons. In other words, the cumulative GDP gain is larger than the underlying cumulative government spending following the impulse in period t. On impact, the fiscal multiplier is already 1.2 and further increases to 3.4 after two years. The multipliers for non-EDP episodes are essentially zero across all horizons.

In the first part of Table 5.2, we report the point estimates of the cumulative multipliers and the associated Driscoll-Kraay standard errors. In addition, we report the first-stage F-statistic (based on the test of Montiel Olea and Pflueger, 2013) which can be used to assess whether our instrumental variable is relevant in each state separately. Indeed, we find that the F-statistic is above the 5% critical value across all horizons and in each state. We therefore conclude that our instrument is relevant. Further, we report the point estimate for the difference and associated p-values which are heteroskedasticity and autocorrelation consistent (Driscoll-Kraay, DK) and which are robust in the presence of weak instruments

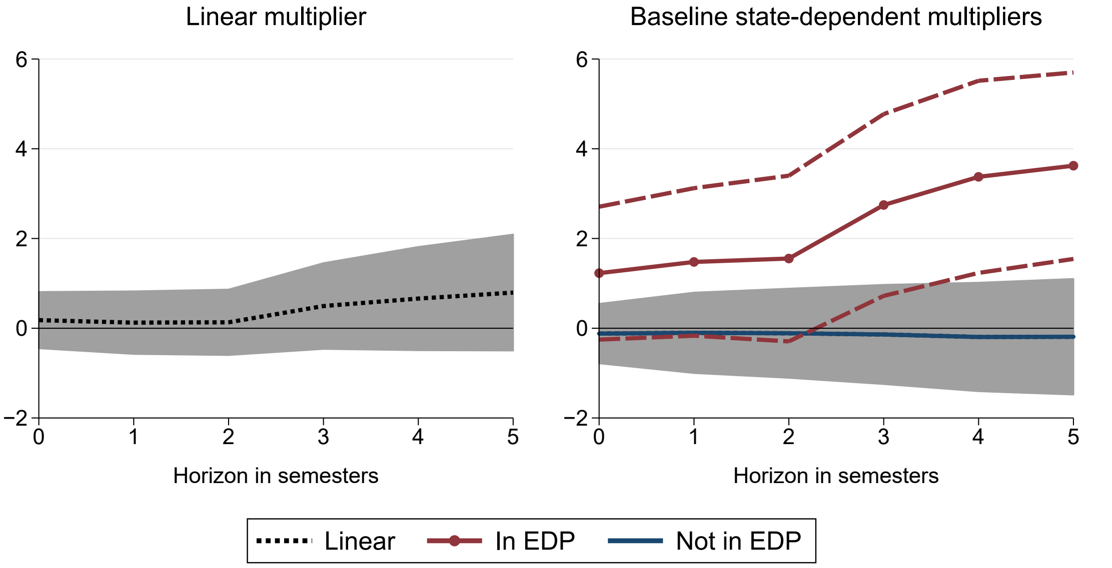

Figure 5.1. Linear and state-dependent cumulative fiscal multipliers

Notes: Estimates in both panels are based on the full sample with 463 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

(Anderson-Rubin, AR).[8] Both p-values confirm our visual observation from Figure 5.1. The difference is statistically significant at the 10% level from horizon three (four) onwards according to the DK (AR) p-values.

These results are not driven by (i) the method we use for the identification of government spending shocks, (ii) the definition of government spending, (iii) the trend GDP specification, (iv) the trend GDP timing, or (v) state-dependent time trends. First, we obtain very similar multipliers if we identify the shocks as originally proposed by Blanchard and Perotti (2002) or if the realization in equation (5.3) is defined as in Auerbach and Gorodnichenko (2013). By using the latter identification approach, we show that the choice of the realization does not influence our results. Second, we confirm the finding of Boehm (2020) who reports that multipliers tend to be larger if one only uses government consumption and disregards government investment. This stresses the importance to incorporate government investment if it is consistently available for a country. Third, we obtain very similar multipliers if we use alternative trend GDP series obtained from polynomials of order 2 or 4, the HP filter using the smoothing parameter $\lambda = 1600$ or the filter proposed by Hamilton (2018).[9] Fourth, multipliers of both states only slightly increase if we scale level variables in equation (5.2) by different lags of trend GDP. This is indirect evidence that our baseline trend specification, a polynomial of order 3, does not vary much with the cycle. Finally, EDP multipliers are large if

8 We implement the AR test following Ramey and Zubairy (2018) and Bernardini and Peersman (2018).

9 Based on Hamilton (2018), we set parameters to $h = 4$ and $p = 2$ (half of the parameters suggested for quarterly data).

Table 5.2—Detailed results for the baseline specification

	Not in EDP			In EDP			Difference		
h	$M_{0,h}$	SE	F-stat.	$M_{1,h}$	SE	F-stat.	P. E.	p (DK)	p (AR)
Baseline (463 observations)									
0	−0.12	(0.40)	209.50	1.23	(0.90)	74.23	1.35	0.20	0.22
1	−0.10	(0.55)	163.86	1.48	(1.00)	61.13	1.58	0.22	0.24
2	−0.11	(0.61)	73.70	1.55	(1.12)	45.36	1.67	0.28	0.30
3	−0.14	(0.67)	54.67	2.75	(1.23)	33.30	2.89	0.08	0.12
4	−0.19	(0.74)	44.26	3.37	(1.30)	26.15	3.57	0.04	0.07
5	−0.19	(0.78)	42.13	3.62	(1.26)	24.10	3.81	0.02	0.05
Strict state definition (286 observations)									
0	0.08	(0.41)	114.64	0.68	(0.53)	133.75	0.60	0.24	0.31
1	0.02	(0.58)	94.87	1.00	(0.68)	125.83	0.99	0.17	0.24
2	−0.06	(0.60)	70.09	1.28	(0.78)	54.72	1.34	0.13	0.19
3	−0.24	(0.65)	51.05	2.78	(1.12)	39.71	3.02	0.01	0.05
4	−0.36	(0.69)	39.07	3.51	(1.19)	29.85	3.87	0.00	0.03
5	−0.32	(0.71)	37.55	3.60	(1.06)	25.86	3.92	0.00	0.02

Notes: We refer to the horizon by h. $M_{\bullet,h}$ denotes the point estimate of the multiplier in the respective state, “SE” the associated Driscoll-Kraay standard error, and “F-stat.” the associated first-stage F-statistic. The critical values for the F-statistic are always 23.1 and 19.7 at the 5% and 10% significance level, respectively. We also report the point estimate of the difference between the two multipliers, “P. E.”, and the associated Driscoll-Kraay (DK) and weak instrument robust Anderson-Rubin (AR) p-values.

time trend coefficients can vary between states. We provide the results for these robustness checks in Appendix 5.A.1.

The baseline multipliers shown in Figure 5.1 are based on our full sample and the estimation does not account for countries entering or leaving the EDP. Thus, the estimation does not necessarily include the same countries across all horizons. This could of course produce misleading estimates. For example, the GDP reaction could be always strongly positive just after the EDP ended. Figure 5.2 underlines that this is not the case. The left panel displays once again the baseline multipliers for comparison purposes. The right panel shows the multipliers for a specification with a strict definition of the state: We restrict the sample to countries remaining in the same state (EDP/non-EDP) across all horizons used in the estimation. Hence, we ensure that we only include countries and episodes for which we have observations in the semesters zero to five. The estimates are based on a smaller sample size, that is, 289 observations at each horizon, but the result remains robust. Multipliers for EDP episodes are positive across all horizons and larger than one after one semester. The second part of Table 5.2 shows that the difference between EDP and non-EDP states becomes even slightly more pronounced and is now statistically significant at the 10% level for $h \geq 3$ according to the DK and AR p-values. The F-statistics for the state-dependent multipliers still show strong evidence that our instrumental variable is relevant.

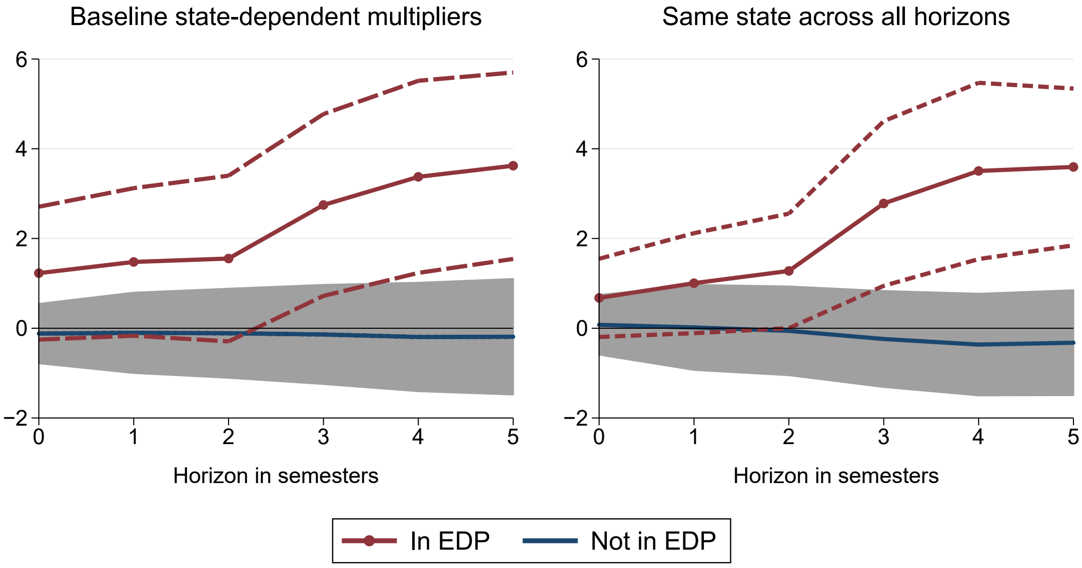

Figure 5.2. Cumulative fiscal multipliers being in the same state across all horizons

Notes: Estimates in the left panel are based on the full sample. Estimates in the right panel are based on 286 observations covering 15 countries. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

One could be concerned that higher state-dependent fiscal spending multipliers in EDP are driven by significantly less government spending in the EDP and/or an asymmetric distribution of government spending shocks across the two states. A simple approach to test the first concern is to regress the share of government spending over GDP on the EDP dummy including country-specific constants. We find little evidence that the share is substantially different in the two states. In addition, we find little evidence that the shock distributions for the two states are a possible driver of the higher multipliers. The mean of the shocks occurring in the EDP (0.19 percentage points) is smaller than the mean of the shocks outside the EDP (0.35 percentage points) while the standard deviation is roughly the same (1.25 in the EDP and 1.18 outside the EDP). Restricting the sample only to the negative (positive) shocks also shows that the shock mean is always smaller in the EDP. Negative shocks occur as often in the EDP (40%) as outside the EDP (38%). Finally, we directly estimate state-dependent multipliers from negative shocks only. The multipliers in EDP are not different from our baseline results, see Appendix 5.A.1. Hence, higher state-dependent fiscal multipliers in EDP are not driven by negative government spending shocks during EDP episodes and a contractionary shock affects the fiscal multipliers like the average shock in our baseline specification.

The results obtained from our baseline specification show that the fiscal multipliers are significantly larger for countries that are in an ongoing EDP. The multipliers in these countries are strictly larger than one, implying that government spending is more effective in the sense that the cumulative GDP gain exceeds the underlying cumulative government spending. We explore the mechanisms behind our results in the next section.

5.4.2 Mechanisms

The size of fiscal spending multipliers is determined by many factors and depends in particular on the dynamics of other macroeconomic variables. Our baseline estimates suggest that government spending is more effective for countries in EDP. In order to rationalize this finding, we investigate the dynamics of the other variables included in our empirical model. Figure 5.3 shows the impulse response functions of these variables to a 1% increase in government spending, as estimated from equation (5.1). The first row displays the responses of government spending, private consumption and private investment (measured in percent). The second row shows the responses of the government bond spread, the marginal tax rate (both measured in percentage points) and the public debt level (measured in percent).

The responses provide evidence that the difference in fiscal multipliers between the EDP and non-EDP samples is mainly driven by diverging investment dynamics. We observe similar investment dynamics in each of our robustness exercises (i)–(v). In the EDP sample, we observe a substantial (and mostly significant) positive response of private investment to a positive government spending shock. Private investment in the non-EDP sample, however, is essentially not reacting to the shock. The path of government spending is similar across states, while we observe a significant positive response of consumption in non-EDP episodes at shorter horizons.[10] The government bond spread (vis-à-vis Germany), as a measure for long-term interest rates, shows a clearly negative reaction for the EDP sample. This is in line with the observed response of private investment: lower interest rates stimulate investment. Bond spreads do not respond to the government spending shock in non-EDP episodes. The initial rise in the marginal tax rate in the EDP sample reflects the reaction of the financing side of the government budget balance to the increase in government spending induced by the shock. Finally, we observe that countries significantly reduce their public debt level in response to the fiscal impulse during EDP episodes. This reaction is in accordance with the EDP's objective of encouraging countries to bring debt levels under control. There is no response of the public debt level in the non-EDP sample. Using the strict state definition from the previous section, we observe even more pronounced reactions of private investment, the bond spread and the public debt level during EDP episodes, see Appendix 5.A.1.

Overall, we find evidence that fiscal multipliers are larger for countries in EDP because government spending provides a stronger stimulus to economic activity. In response to a positive spending shock, countries in EDP achieve a significant reduction of public debt

[10] We do not observe government spending reversals, which is consistent with the reaction of the debt level in our sample, see Corsetti et al. (2012a).

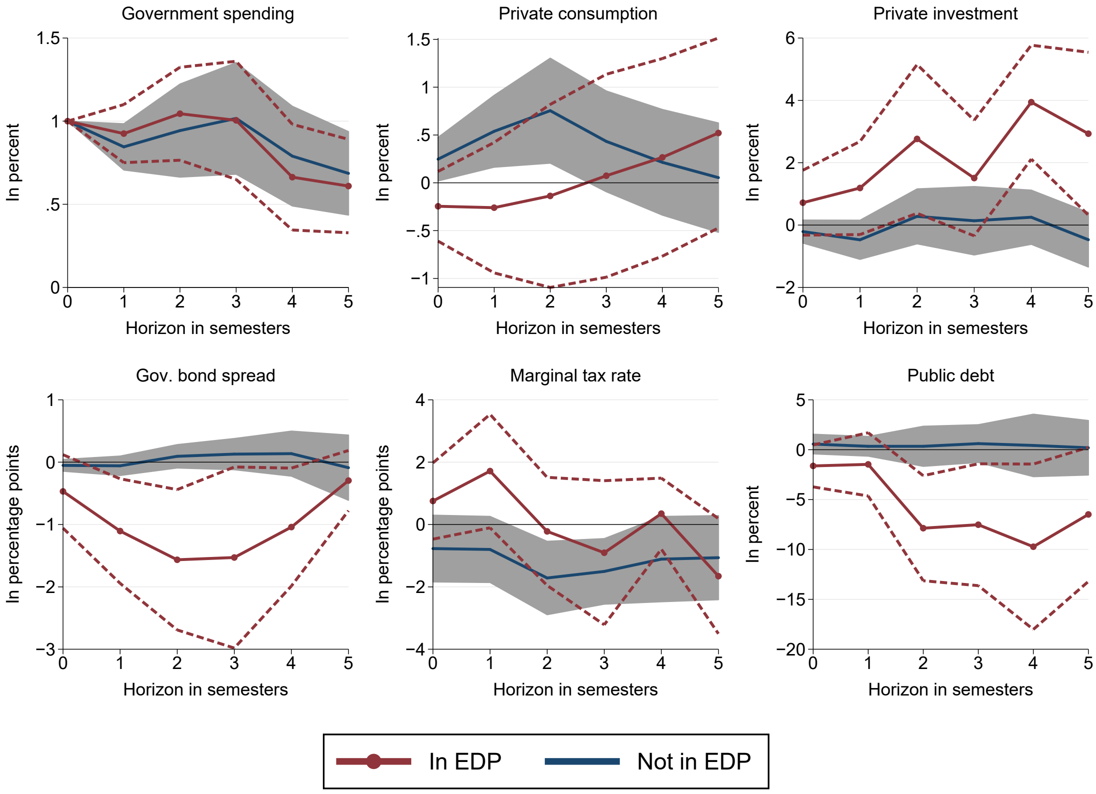

Figure 5.3. Impulse response functions

Notes: Estimates in each panel are based on the full sample. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

and long-term interest rates decrease, signaling a stable fiscal outlook. This boosts private investment and gives rise to substantial crowding-in. One possible explanation is that being in the EDP demonstrates credible commitment to fiscal discipline and therefore leads to lower risk premia. Thus, the EDP fulfills its task as corrective arm of the EU fiscal framework, while at the same time ensuring the effectiveness of fiscal stimulus.

5.4.3 Fiscal Multipliers in Bad Times

The sample used in our baseline estimation includes a broad set of countries which might differ in many dimensions. For example, countries have varying fiscal positions or face different cyclical fluctuations over time. The EDP is designed to ensure stable public finances, which is more likely to be an issue in member countries going through bad times. Thus, the measures implemented by the EDP are supposed to be especially effective for these countries. Economic theory provides different explanations for the potentially different size of the fiscal multiplier in bad times. From a Keynesian point of view, multipliers are larger during times of slack because government spending is less likely to crowd out consumption

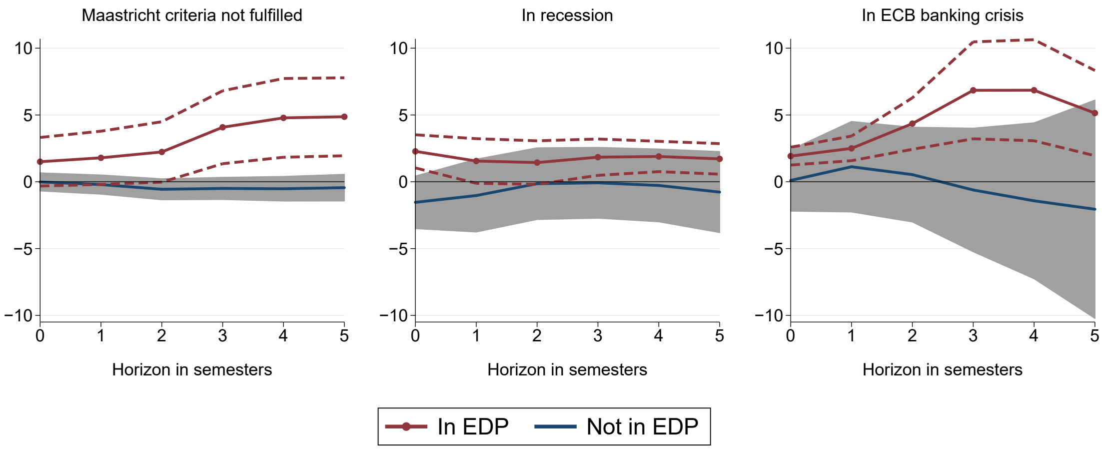

Figure 5.4. Cumulative fiscal multipliers in bad times

Notes: Estimates are based on the following number of observations and countries. Left: 249 observations and 14 countries. Middle: 92 observations and 15 countries. Right: 122 observations and 14 countries. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

and/or investment. From a neoclassical perspective, however, consumption decreases when government spending (and deficits) increase significantly since consumption depends on intertemporal optimization. In order to isolate the effect of the EDP on the multiplier in bad times, we estimate equation (5.2) for different subsamples. For all subsamples, the relevant state variable is the EDP dummy.

We show the cumulative fiscal multipliers in Figure 5.4 and report detailed test results in Table 5.3. Supporting impulse responses can be found in Appendix 5.A.2. The left panel of Figure 5.4 considers a subsample with episodes of countries which do not fulfill the Maastricht criteria for either deficit only (15% of the observations) or debt only (68%) or both (17%). These countries have in common that they are in a weak fiscal position with a special emphasis on unsustainable debt levels as the debt exceeds the Maastricht criterion for 85% of the observations. The estimates of the fiscal multipliers indicate that the EDP multipliers are statistically different from the non-EDP multipliers at the 10% level for $h > 1$. This shows that the EDP indeed increases the effectiveness of government spending in this subsample. Multipliers for the EDP sample are even larger than in the specification using the full sample as shown in the right panel of Figure 5.1. This is due to a more pronounced reaction of interest rates and private investment. The first-stage F-statistic exceeds the 5% critical value at all horizons in the non-EDP state and the 10% critical value for $h \leq 4$ in the EDP state, suggesting that our results do not suffer from a weakly identified instrument. The results are similar if we condition the subsample only on the observations for which debt exceeds the Maastricht criterion.

Table 5.3—Detailed results for multipliers in bad times

	Not in EDP			In EDP			Difference		
h	$M_{0,h}$	SE	F-stat.	$M_{1,h}$	SE	F-stat.	P. E.	p-val (DK)	p-val (AR)
Maastricht criteria not fulfilled (249 observations)									
0	−0.01	(0.41)	375.84	1.50	(1.11)	58.97	1.51	0.19	0.20
1	−0.21	(0.43)	251.18	1.79	(1.21)	64.34	2.00	0.11	0.12
2	−0.56	(0.48)	74.40	2.23	(1.37)	42.58	2.80	0.05	0.05
3	−0.50	(0.50)	56.55	4.08	(1.66)	30.96	4.58	0.01	0.01
4	−0.52	(0.56)	43.84	4.78	(1.79)	22.74	5.31	0.00	0.01
5	−0.44	(0.60)	36.69	4.87	(1.77)	19.14	5.31	0.00	0.00
Debt criterion not fulfilled (210 observations)									
0	0.79	(0.60)	135.81	1.66	(0.96)	114.82	0.87	0.46	0.45
1	0.68	(0.72)	107.29	2.44	(1.27)	64.44	1.75	0.21	0.22
2	0.14	(0.90)	66.29	3.49	(1.70)	23.54	3.35	0.05	0.06
3	0.35	(1.11)	43.32	5.88	(2.26)	13.24	5.53	0.01	0.02
4	0.33	(1.20)	28.59	6.09	(2.33)	9.61	5.76	0.00	0.01
5	0.28	(1.15)	23.25	5.14	(1.87)	8.27	4.86	0.00	0.00
In recession (92 observations)									
0	−1.54	(1.19)	50.98	2.28	(0.75)	34.06	3.82	0.02	0.15
1	−1.03	(1.66)	41.95	1.55	(1.02)	22.55	2.59	0.25	0.33
2	−0.14	(1.63)	50.32	1.44	(0.99)	13.47	1.58	0.46	0.47
3	−0.08	(1.61)	28.28	1.84	(0.83)	9.76	1.92	0.31	0.31
4	−0.27	(1.65)	12.30	1.89	(0.69)	8.03	2.17	0.20	0.18
5	−0.77	(1.84)	6.03	1.71	(0.70)	7.52	2.48	0.13	0.11
In ECB banking crisis (122 observations)									
0	0.10	(1.40)	96.05	1.92	(0.41)	46.85	1.82	0.19	0.24
1	1.13	(2.06)	81.63	2.50	(0.56)	30.32	1.37	0.50	0.51
2	0.54	(2.15)	20.97	4.35	(1.17)	12.92	3.82	0.12	0.15
3	−0.61	(2.81)	13.93	6.85	(2.21)	11.16	7.46	0.06	0.07
4	−1.42	(3.55)	6.27	6.85	(2.30)	9.97	8.28	0.06	0.06
5	−2.06	(4.97)	2.00	5.14	(1.94)	9.38	7.20	0.14	0.10

Notes: We refer to the horizon by h. $M_{\bullet,h}$ denotes the point estimate of the multiplier in the respective state, "SE" the associated Driscoll-Kraay standard error, and "F-stat." the associated first-stage F-statistic. The critical values for the F-statistic are always 23.1 and 19.7 at the 5% and 10% significance level, respectively. We also report the point estimate of the difference between the two multipliers, "P. E.", and the associated Driscoll-Kraay (DK) and weak instrument robust Anderson-Rubin (AR) p-values.

The middle panel focuses on a subsample with recessionary episodes. We identify recessions using the simple and transparent algorithm proposed by Harding and Pagan (2002) based on quarterly real GDP growth from the OECD Main Economic Indicators 2020–01.[11] The EDP multipliers are significantly positive for countries in a recession, but considerably smaller than the baseline multipliers at longer horizons. This can be rationalized by a more persistent increase in government spending in response to the shock, along with a crowding-in of consumption and a slight crowding-out of private investment in this subsample. Note that the

[11] As originally suggested by Harding and Pagan (2002), we require that complete cycles have a length of at least five quarters and that a cycle phase lasts at least two quarters. We list the identified recessions which are included in the baseline sample in Appendix 5.A.3.

EDP/non-EDP multipliers are not statistically different and the first-stage F-statistic suggests that the instrumental variable is less relevant at longer horizons.

Finally, the right panel shows the multipliers for periods in which countries are hit by a banking crisis, as defined in the European Financial Crises Database provided by the European Central Bank and the European Systemic Risk Board (Lo Duca et al., 2017).[12] Fiscal multipliers in the EDP sample are significantly positive and even larger than in the baseline estimation at longer horizons due to strong crowding-in of private investment. At the same time, the instrumental variable is less relevant for these horizons. Multipliers for the non-EDP sample are not significantly different from zero across all horizons. However, the difference between the two states is only statistically significant at the 10% level for horizons three and four.

The cumulative fiscal multiplier estimates for these subsamples confirm that the EDP is successful in increasing effectiveness of government spending for countries in bad times. In particular, the EDP seems to be fully functional for countries with a weak fiscal position (as indicated by non-compliance with the Maastricht criteria). This indicates that the procedure fulfills the purpose it was designed for.

5.5 Alternative State Variables

We have so far implicitly assumed that the indicator whether a member state is in the EDP or not has explanatory power for the differences in fiscal multipliers across countries. However, countries in the EDP could tend to have weak fiscal positions or to experience recessions more frequently than other countries, both of which could in turn explain the variations in fiscal multipliers. Potentially, the EDP/non-EDP states could simply be a proxy for different underlying state variables. In Figure 5.5, we present fiscal multipliers for various alternative state variables. That is, we re-estimate equation (5.2) using different indicator variables $\mathcal{I}_{i,t-1}$. The estimates refer to cumulative two-year multipliers (horizon $h = 4$) with the corresponding 90% confidence intervals. For comparison purposes, the first row shows the two-year multipliers from our baseline specification. The other rows report the multipliers if different combinations of non-compliance with the Maastricht criteria, recessions or banking crises are used as state variables. The recession and banking crisis dummy variables are constructed as described in the previous section. We provide corresponding impulse response functions in Appendix 5.A.4.

Again, non-compliance with the Maastricht criteria is interpreted as a signal for a weak fiscal position. We observe that the corresponding multipliers are smaller than the EDP

[12] The most recent version of the database was published in 2017. We extend the data using the warnings issued by the European Systemic Risk Board. In fact, the European Systemic Risk Board did not identify any banking crises in 2018 and the first half of 2019.

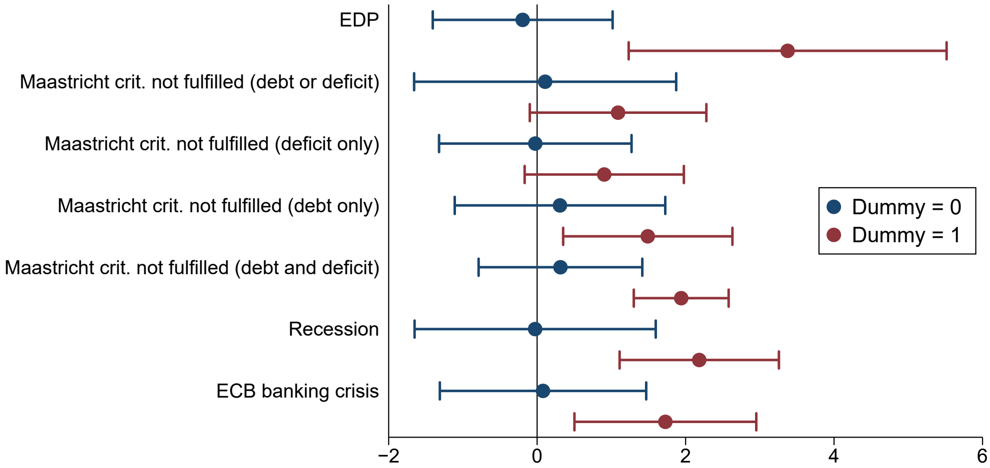

Figure 5.5. Cumulative two-year fiscal multipliers ($h = 4$)

Notes: Full sample for each state dependency. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

multipliers, but still positive. This stands in contrast to the findings reported in the literature. Corsetti et al. (2012b), Nickel and Tudyka (2014) and Banerjee and Zampolli (2019) find that crowding-out of private investment leads to lower and even negative multipliers in countries with weak fiscal positions, while Huidrom et al. (2020) explains the lower multipliers in these countries by a decrease in consumption. In this context, Banerjee and Zampolli (2019) and Huidrom et al. (2020) document the relevance of the transmission via the interest rate channel: Interest rates in high-debt countries rise in response to fiscal stimulus, reducing investment and consumption. By contrast, we observe a decrease in interest rates after a positive government spending shock for non-compliers with the Maastricht criteria which leads to crowding-in of private investment, thereby rationalizing positive multipliers.

Our results from Section 5.4.3 suggest that the EDP can explain these dynamics. In fact, Figure 5.4 and corresponding impulse responses show that the multipliers are positive, interest rates are lower and investment is crowded in only if a country with a weak fiscal position is in the EDP. However, for countries with weak fiscal positions outside the EDP, the multipliers are close to zero and private investment is crowded out due to a slight increase in interest rates. One possible explanation is that being in the EDP is a positive signal to investors that sovereign credit risk is reduced. This would lower risk premia and interest rates and in turn boost investment.

In line with Auerbach and Gorodnichenko (2012, 2013), we find significantly positive multipliers in recessions driven by crowding-in of private consumption and investment. We also find positive multipliers during banking crises. As in Corsetti et al. (2012b), we observe a positive reaction of investment for these episodes. Note that the multipliers associated with these states are smaller than the EDP multipliers.

The fiscal multipliers for all these alternative states are positive, suggesting that these state variables play a role. However, the multipliers are not significantly different from the respective other state (except for our baseline, the EDP state variable) and the estimated magnitudes are at most half the size of the EDP multiplier. These results confirm that the EDP is not only a proxy for different underlying factors. The EDP seems to be a suitable state variable for explaining variations in fiscal multipliers.

5.6 Fiscal Multipliers in Real Time

Using the rich availability of forecasts in different EO editions, we want to explore in this section how fiscal spending multipliers are observed in real time. This is relevant because policy makers may plan government spending based on the size of future fiscal spending multipliers implied by the projected paths of real GDP and real government spending at a specific time. For example, policy makers expect that an additional unit of government spending pays off more if the forecasts of the multipliers are large.

There is little evidence on how state-dependent fiscal multipliers are observed in real time. Blanchard and Leigh (2013, 2014) and Górnicka et al. (2020) investigate fiscal multipliers for European countries in the period of the European sovereign debt crisis. Both find that multipliers tend to be larger than initially forecasted which they attribute to the learning of the forecasters during the crisis. While Blanchard and Leigh (2013, 2014) provide evidence that multipliers exceed one, Górnicka et al. (2020) cannot confirm this finding. To our knowledge, we are the first to present evidence on how fiscal multipliers are observed in real time in different states. In particular, we can thereby investigate whether forecasters learn in a similar fashion in each state.

We use a version of equation (5.2) in which we estimate real-time multipliers $M^s_{0,h}$ and $M^s_{1,h}$ by including real-time data from EO edition s:

$$\begin{aligned}\sum_{j=0}^{h} Y^s_{i,t+j} \;=\; & \left(1-\mathcal{I}_{i,t-1}\right)\left[\alpha^s_{0,i,h} + M^s_{0,h}\sum_{j=0}^{h} G^s_{i,t+j} + \Phi^s_{0,h}(L)X^s_{i,t-1}\right] \\ & + \mathcal{I}_{i,t-1}\left[\alpha^s_{1,i,h} + M^s_{1,h}\sum_{j=0}^{h} G^s_{i,t+j} + \Phi^s_{1,h}(L)X^s_{i,t-1}\right] + \sum_{k=1}^{2}\psi^s_k T^k_t + \varepsilon^s_{i,t+h}\,.\end{aligned} \tag{5.4}$$

Our approach features three distinct dimensions which explicitly account for the available information set of the forecasters at time s. First, the variables $Y^s_{i,t}$ and $G^s_{i,t}$ are now forecasts (nowcasts) if $t > s$ ($t = s$). Variables, for which $t < s$ is true, are ex-post values. Second, the instrument should contain only the information available up to time s, too. We therefore compute the forecast error in equation (5.3) by using the realized growth rate as of time

s, $g_{i,t}^{s}$, instead of the realized growth rate reported in EO edition 106, $g_{i,t}^{2019H2}$. Hence, the forecast error comparing the forecast and the realization is well defined for $s = t + k$, with $k \geq 1$:

$$FE_{i,t}^{s} = g_{i,t}^{s} - g_{i,t}^{t-1} \,. \tag{5.5}$$

Third, several studies show that trend estimates are unreliable in real time (e.g., Orphanides and van Norden, 2002; Orphanides, 2003). Given this uncertainty around the trend estimates, especially at the sample end, it is sensitive to scale level variables by a lag of the estimated trend rather than by the contemporaneous estimated trend.[13]

Figure 5.6 shows the real-time fiscal multipliers for $s = t + 1$ (first column) and $s = t + 5$ (second column). We further include the impulse responses of government spending and GDP for each specification. Our choice of s implies that the vector of control variables $X_{i,t-1}^{s}$ always contains past values.[14] Multipliers to the left of the dashed line in the first column are based on past values, while the multipliers to the right of the dashed line are partially based on nowcasts ($h = 1$) or on nowcasts and forecasts ($h \geq 2$). In the right column, multipliers are entirely based on ex-post values. We report corresponding statistics in Table 5.4 and supporting impulse responses can be found in Appendix 5.A.5.

Fiscal multipliers in the upper right panel are similar to the multipliers from Figure 5.1 because both rely on ex-post values. One possible reason for the smaller EDP multipliers in Figure 5.6 is that the levels of GDP and government spending at $h = 4$ are still based on the first release and can be subject to further revisions. The EDP multipliers are still larger than one and significantly positive after one semester. The non-EDP multipliers are smaller than one and not significantly different from zero again. In the upper left panel, fiscal multipliers at horizons $h \geq 1$ depend on nowcasts and forecasts for GDP and government spending. The state-dependent multipliers are not very different from each other, confidence intervals overlap and both are not significantly different from zero. The two very distinct EDP multipliers in the first row indicate that fiscal multipliers in the EDP are underestimated in real time. Hence, fiscal stimulus is expected to be less effective than it turns out ex post. The multipliers outside the EDP, however, do not depend on the time of the estimation.

With respect to the impulse responses of government spending and GDP in the EDP, we can trace the underestimation of the multipliers to a lower-than-expected government spending path and a larger-than-expected GDP path. The revisions of the GDP path seem larger than

[13] This approach is valid as long as the trend does not fluctuate with the business cycle. Indeed, our robustness checks in Appendix 5.A.1 document that our baseline results from Figure 5.1 remain if we re-estimate the multipliers with level variables which are scaled now by different lags of the estimated trend.

[14] As the marginal tax rate is not available in real time, we include observations from EO edition 106 and thereby ignore possible revisions of past values. Given the nature of the variable, we believe that these revisions are negligible.

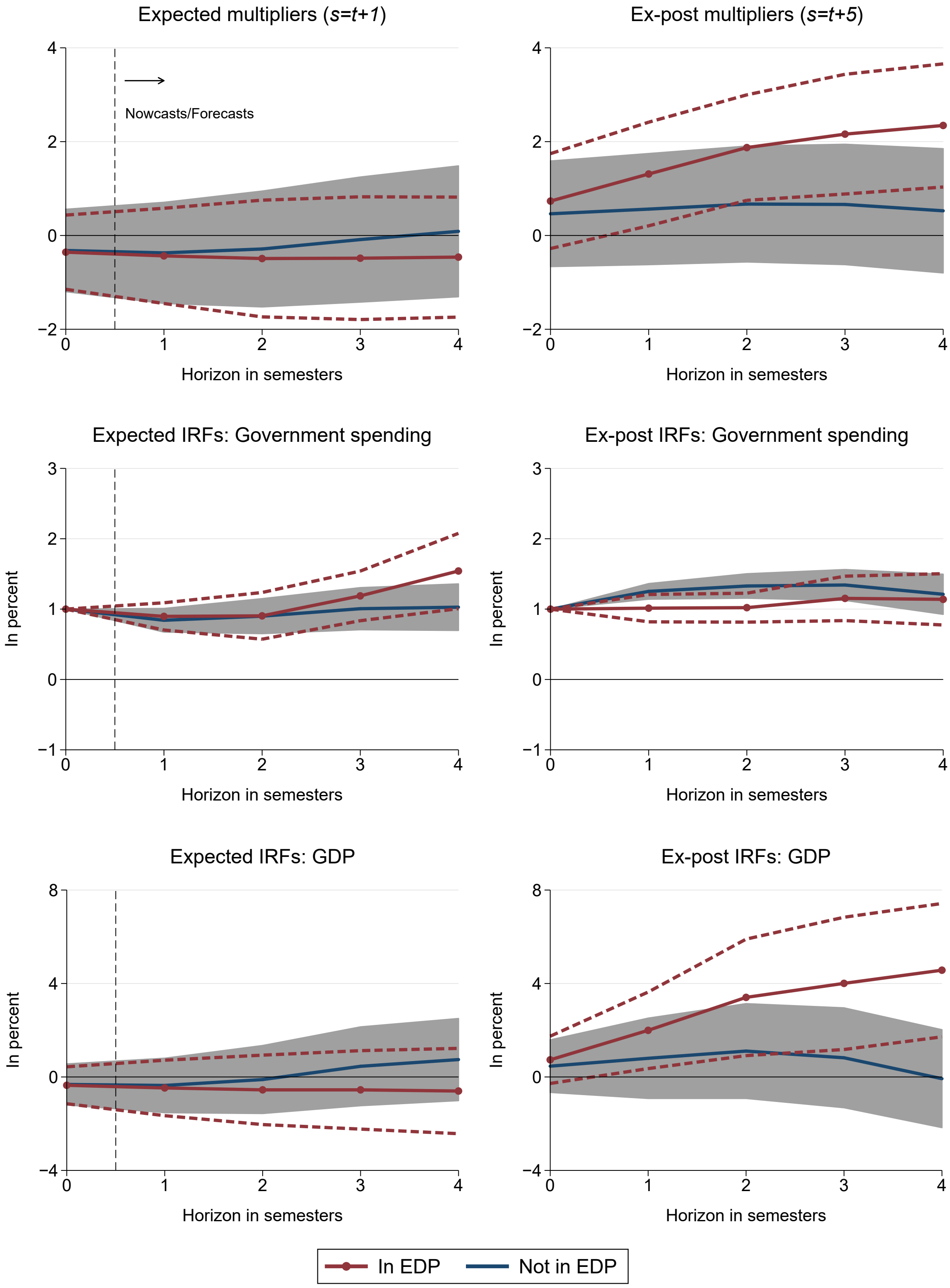

Figure 5.6. Multipliers and impulse responses in real time

Notes: Estimates in each panel are based on 373 observations covering 15 countries. In the left column, multipliers and impulse response functions (IRFs) to the left (right) of the vertical dashed line depend on ex-post real-time data (nowcasts/forecasts). In the right column, multipliers and IRFs always depend on ex-post real-time data. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

Table 5.4—Detailed results for multipliers in real time

	Not in EDP			In EDP			Difference		
h	$M^s_{0,h}$	SE	F-stat.	$M^s_{1,h}$	SE	F-stat.	P. E.	p-val (DK)	p-val (AR)
Expected multipliers for $s = t + 1$									
0	−0.32	(0.53)	70.73	−0.36	(0.48)	112.26	−0.04	0.96	0.96
1	−0.37	(0.66)	40.05	−0.43	(0.62)	47.41	−0.06	0.94	0.94
2	−0.29	(0.75)	26.46	−0.49	(0.76)	26.79	−0.20	0.83	0.83
3	−0.09	(0.81)	21.23	−0.48	(0.80)	24.17	−0.40	0.70	0.70
4	0.09	(0.85)	18.58	−0.46	(0.78)	23.91	−0.55	0.60	0.60
Ex-post multipliers for $s = t + 5$									
0	0.46	(0.68)	358.18	0.73	(0.62)	114.06	0.27	0.80	0.80
1	0.56	(0.72)	285.62	1.31	(0.67)	55.67	0.75	0.48	0.50
2	0.67	(0.75)	200.73	1.87	(0.68)	48.25	1.20	0.28	0.29
3	0.66	(0.78)	146.49	2.16	(0.78)	39.65	1.50	0.21	0.21
4	0.53	(0.81)	104.98	2.35	(0.80)	33.66	1.82	0.13	0.13

Notes: We refer to the horizon by h. $M^s_{\bullet,h}$ denotes the point estimate of the multiplier in the respective state, "SE" the associated Driscoll-Kraay standard error, and "F-stat." the associated first-stage F-statistic. The critical values for the F-statistic are always 23.1 and 19.7 at the 5% and 10% significance level, respectively. We also report the point estimate of the difference between the two multipliers, "P. E.", and the associated Driscoll-Kraay (DK) and weak instrument robust Anderson-Rubin (AR) p-values.

the revisions of the government spending path between the two specifications. Additionally, higher-than-expected private consumption and investment contribute to the change of the GDP response. Impulse responses for the interest spread are not significant but indicate that the spread is ex post lower than expected and therefore boosts private investment.

5.7 Conclusion

We estimate state-dependent fiscal spending multipliers to evaluate the effect of the EDP for 17 EU countries between 2000 and 2019. We show that fiscal multipliers in the EDP are larger than one and significantly different from the multipliers outside the EDP. The analysis of the underlying mechanisms shows that the higher multipliers are mainly driven by the crowding-in of investment which goes along with a significant reduction of public debt and a decrease of long-term interest rates in response to a positive government spending shock. The latter two reactions signal a stable fiscal outlook. Thus, the EDP fulfills its task as corrective arm of the EU fiscal framework, while at the same time ensuring the effectiveness of fiscal stimulus. In addition, we find that the EDP is especially successful in bad times. Furthermore, we show that it is not just a proxy for other underlying factors. Finally, we provide evidence that forecasters underestimate fiscal multipliers in the EDP in real time.

Our results have important policy implications. First, the EDP fulfills the function it was designed for. The output response to government spending is stronger for countries in the procedure. Second, the large EDP multipliers for a country in a weak fiscal position show that the EDP is more effective in bad times. Third, the underestimation of fiscal multipliers in real time masks the ex-post effect of fiscal stimulus. This could mislead policy makers who expect that a change in government spending does not have a substantial effect on the economy.

5.A Appendix

5.A.1 Robustness of Results

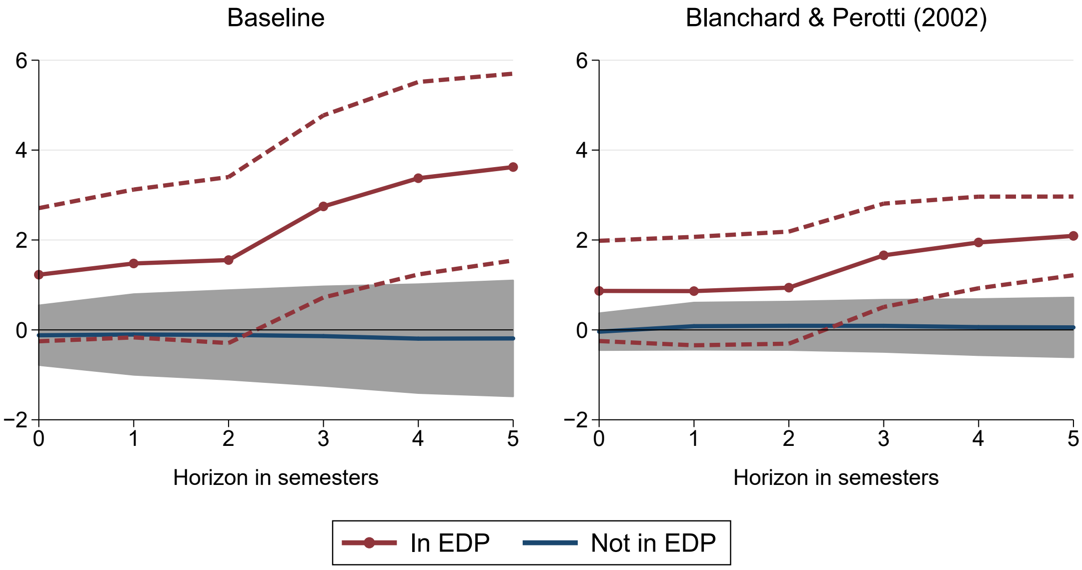

Figure 5.7. Comparison of multipliers obtained from our baseline identification and from the identification of Blanchard and Perotti (2002)

Notes: The left panel repeats the multipliers from Figure 5.1. The right panel shows the multipliers using the shock identification of Blanchard and Perotti (2002). Estimates in both panels are based on the full sample with 463 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

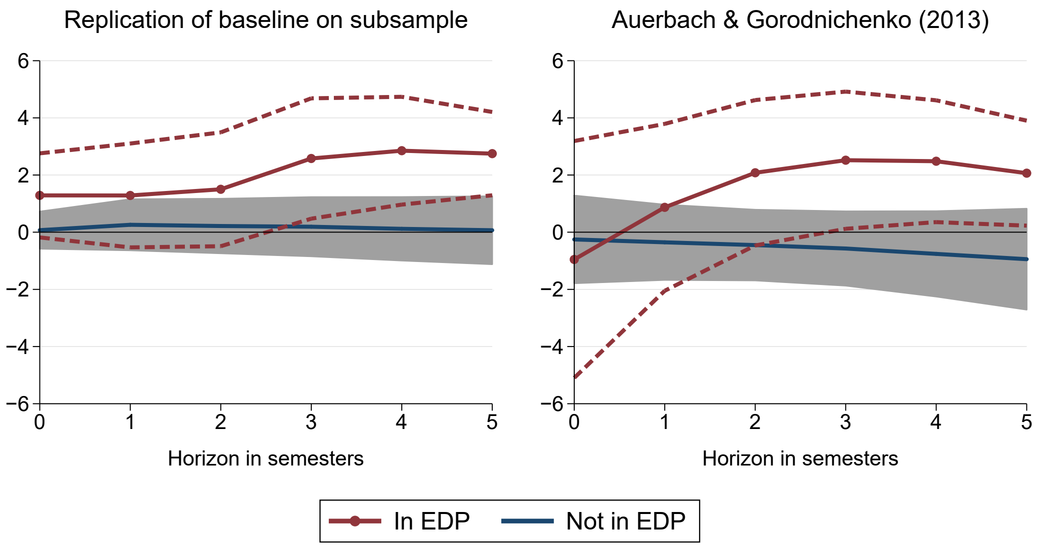

Figure 5.8. Comparison of multipliers obtained from our baseline identification and from the identification of Auerbach and Gorodnichenko (2013)

Notes: The left panel displays the multipliers from Figure 5.1 based on the subsample which the left and right panel have in common. The right panel shows the multipliers using the shock identification of Auerbach and Gorodnichenko (2013). Estimates in both panels are based on 406 observations covering 15 countries. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

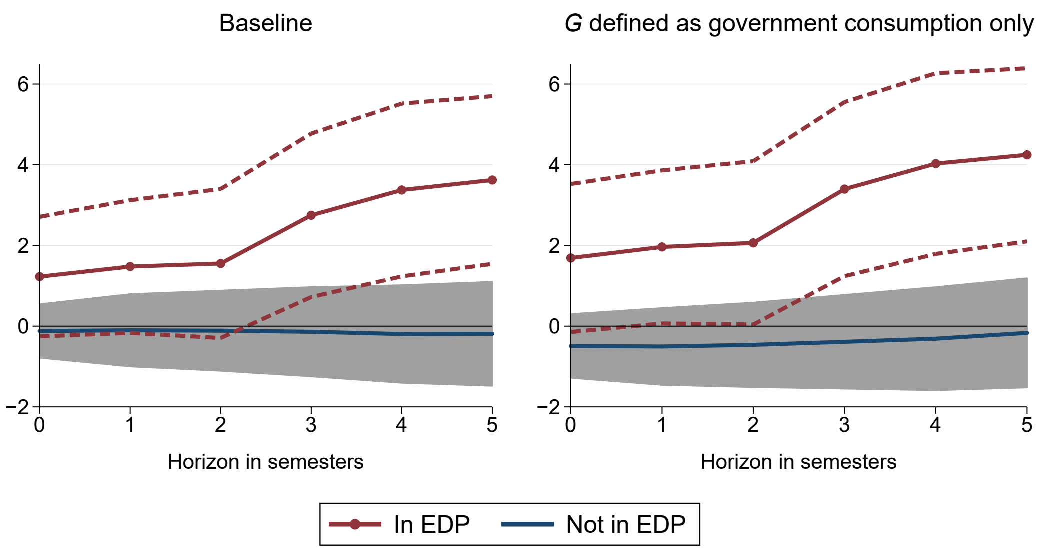

Figure 5.9. Comparison of multipliers obtained from our baseline identification and from the definition of G as government consumption only

Notes: The left panel repeats the multipliers from Figure 5.1. The right panel shows the multipliers using G defined as government consumption only. Estimates in both panels are based on the full sample with 463 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

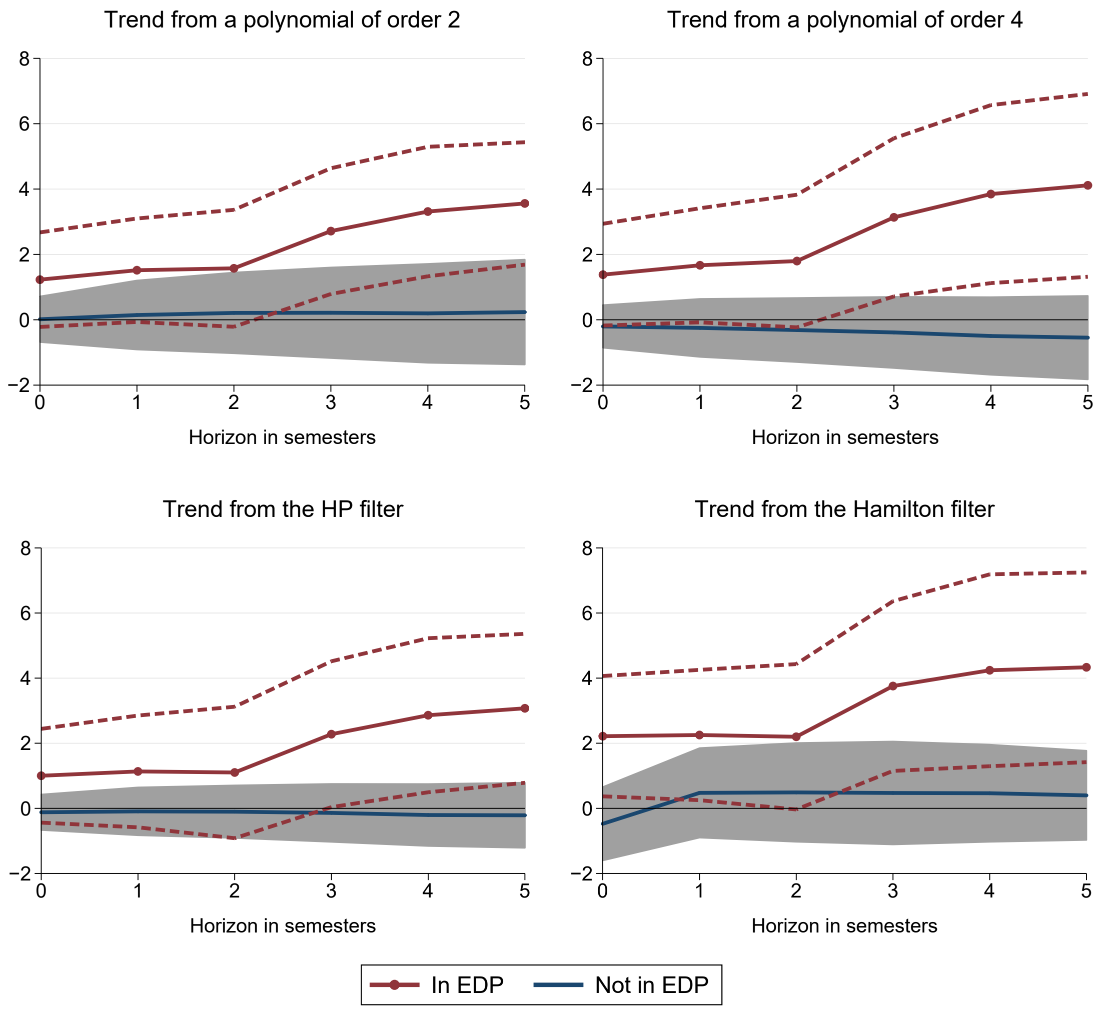

Figure 5.10. Comparison of multipliers using different trend specifications

Notes: Each panel shows multipliers using different specifications to estimate trend GDP. Estimates in all panels are based on the full sample with 463 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

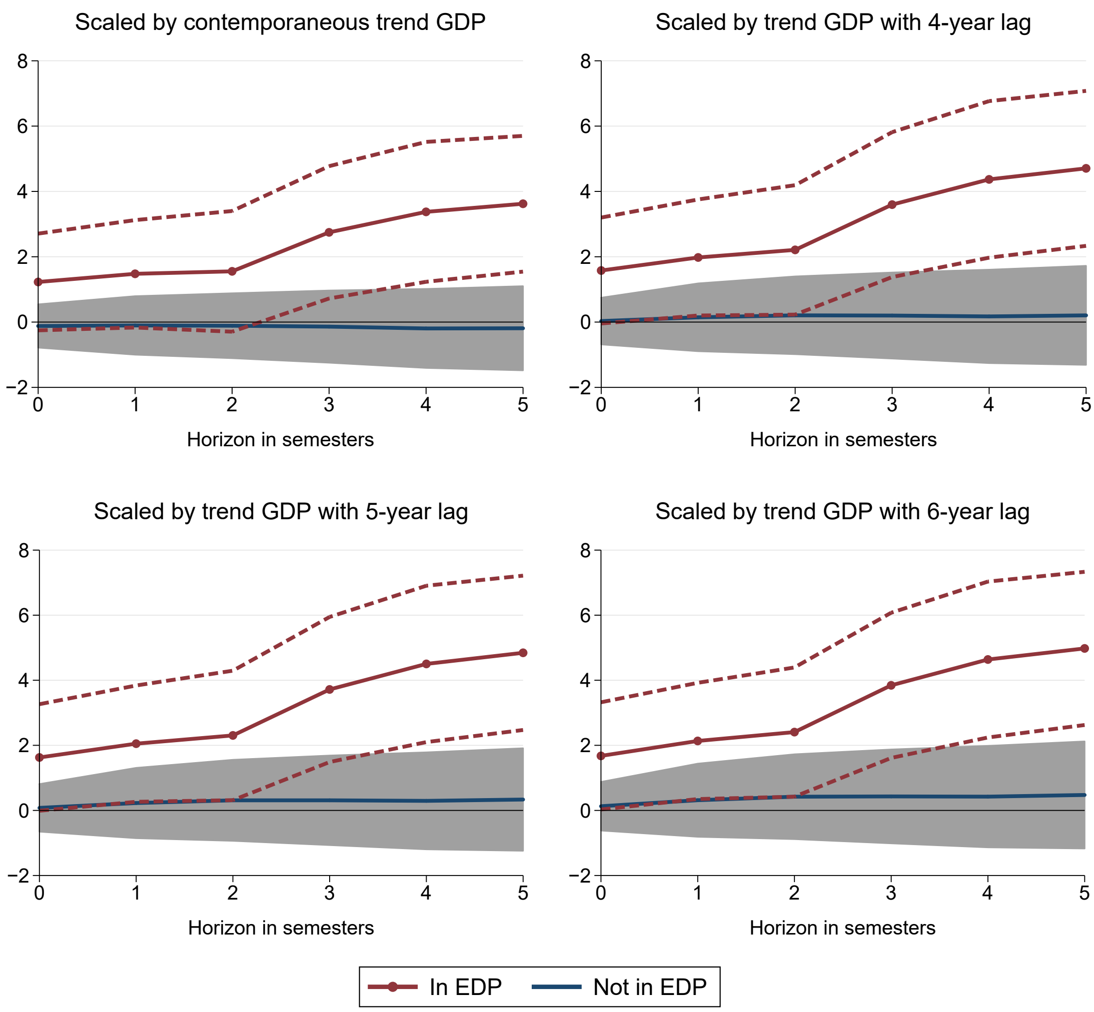

Figure 5.11. Comparison of multipliers using different lags of trend GDP

Notes: The upper left panel repeats the multipliers from Figure 5.1. The other panels show the multipliers using different lags of trend GDP to scale level variables. Estimates in all panels are based on the full sample with 463 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

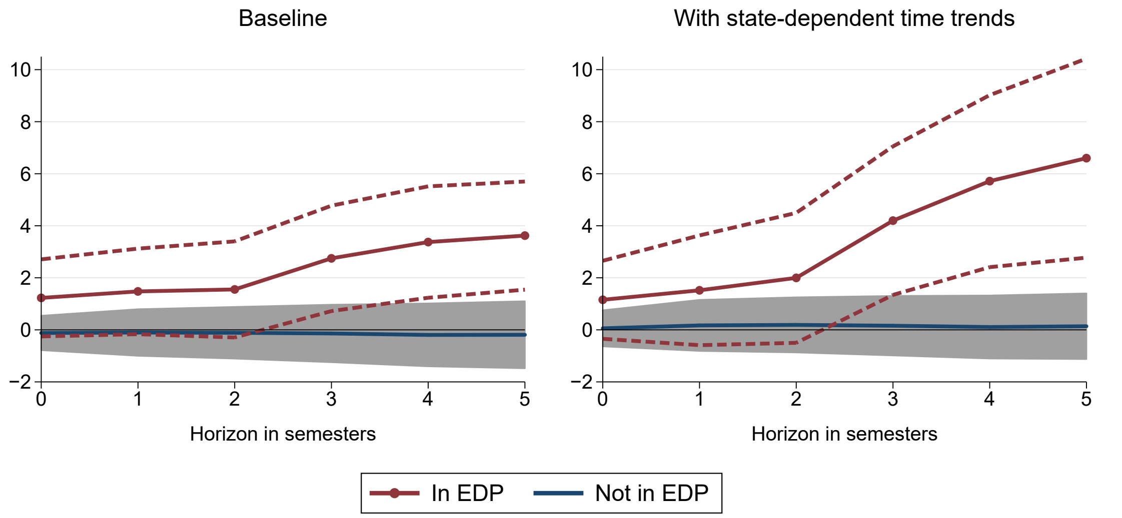

Figure 5.12. Comparison of multipliers obtained from our baseline identification and an identification with state-dependent time trends

Notes: The left panel repeats the multipliers from Figure 5.1. The right panel shows the multipliers with state-dependent time trends. Estimates in both panels are based on the full sample with 463 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

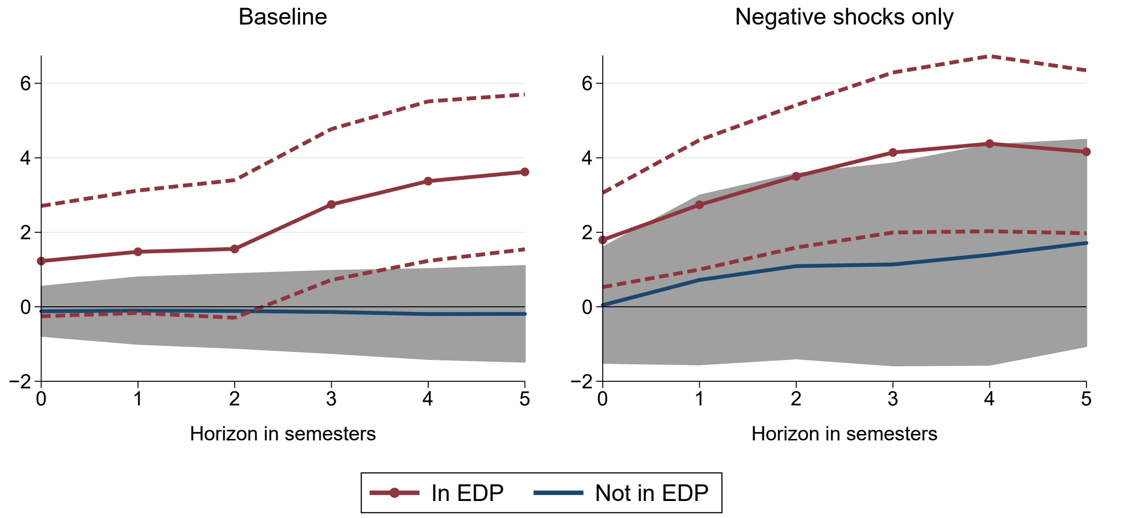

Figure 5.13. Comparison of multipliers obtained from our baseline identification and from negative shocks only

Notes: The left panel repeats the multipliers from Figure 5.1. The right panel shows the multipliers from negative shocks only. Estimates in the left (right) panel are based on 463 (180) observations covering 17 countries. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

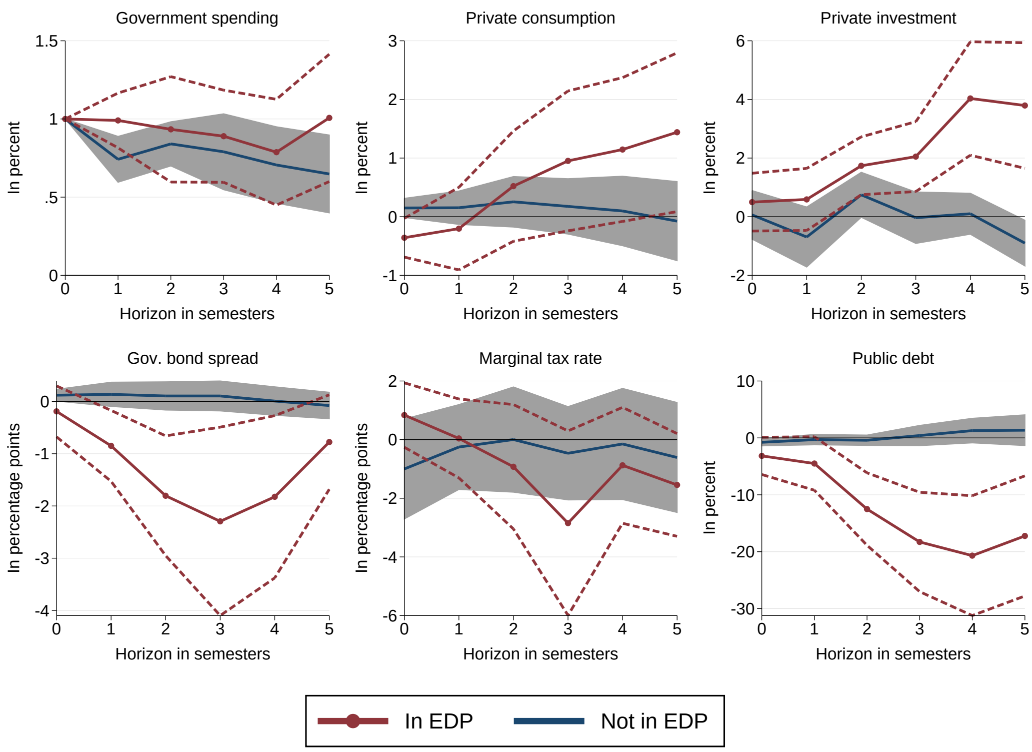

Figure 5.14. IRFs for strict-state definition

Notes: Estimates in each panel are based on 286 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

5.A.2 Detailed Results for Multipliers in Bad Times

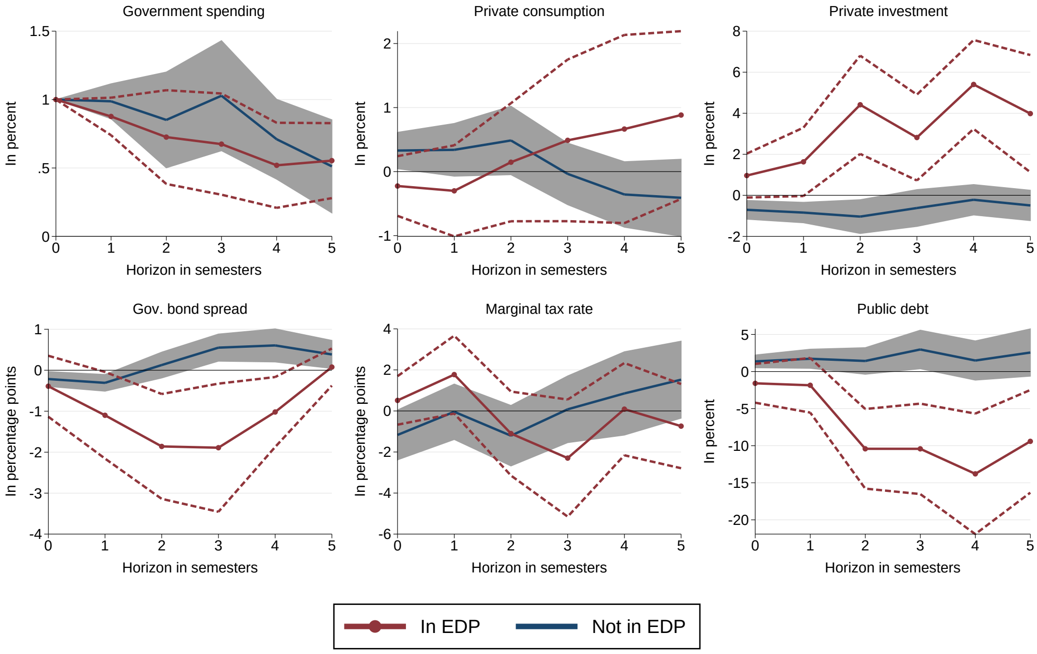

Figure 5.15. IRFs in bad times – Maastricht criteria not fulfilled (debt or deficit)

Notes: Estimates in each panel are based on 249 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

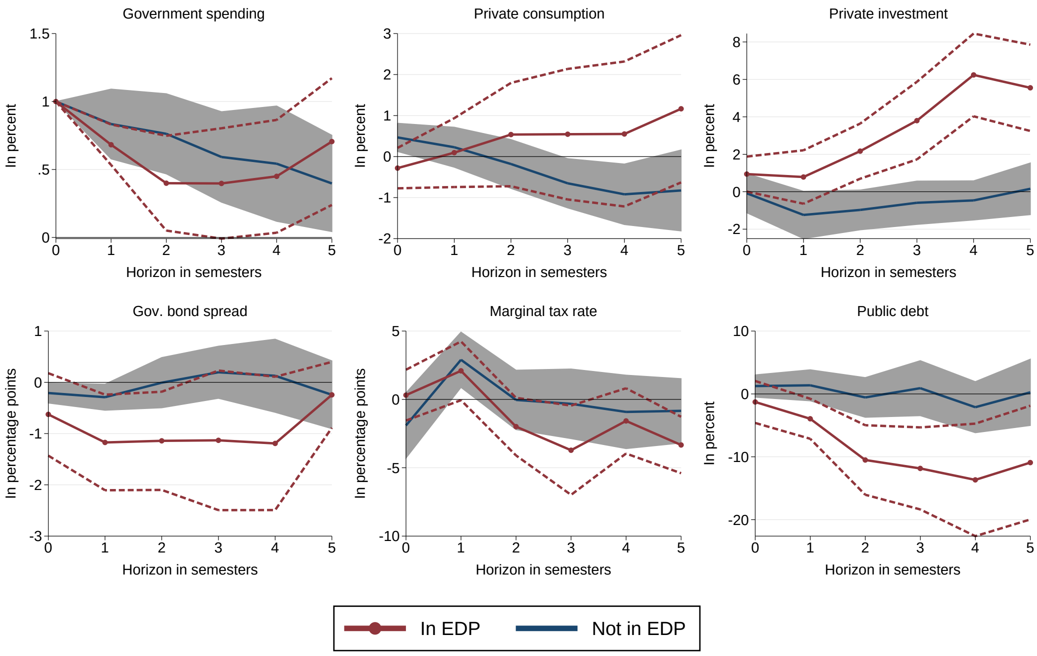

Figure 5.16. IRFs in bad times – Debt criterion not fulfilled

Notes: Estimates in each panel are based on 210 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

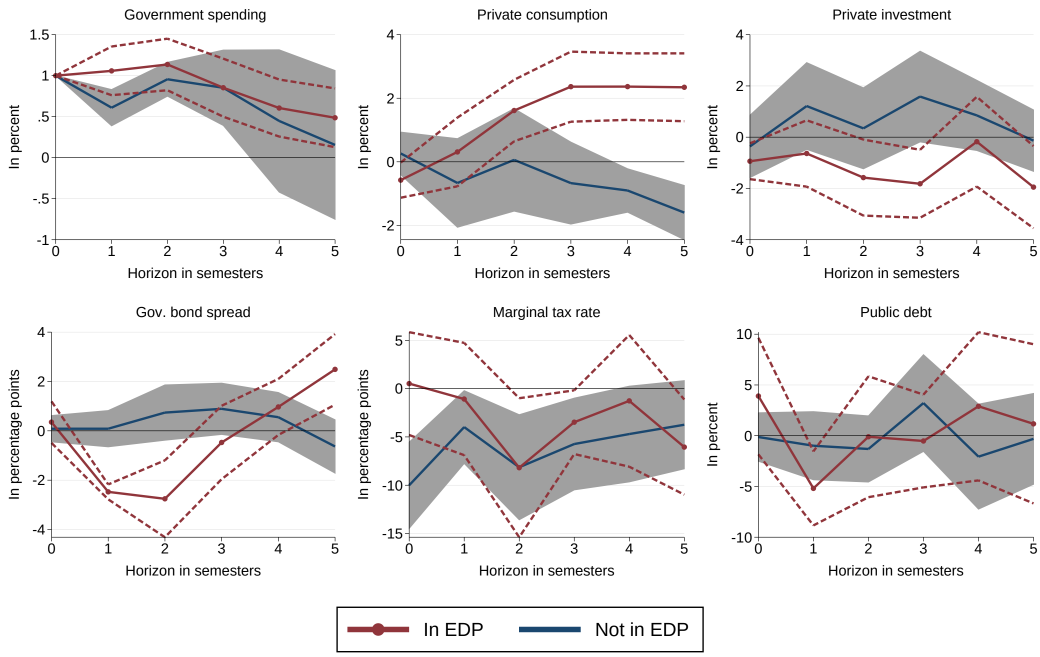

Figure 5.17. IRFs in bad times – In recession

Notes: Estimates in each panel are based on 92 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

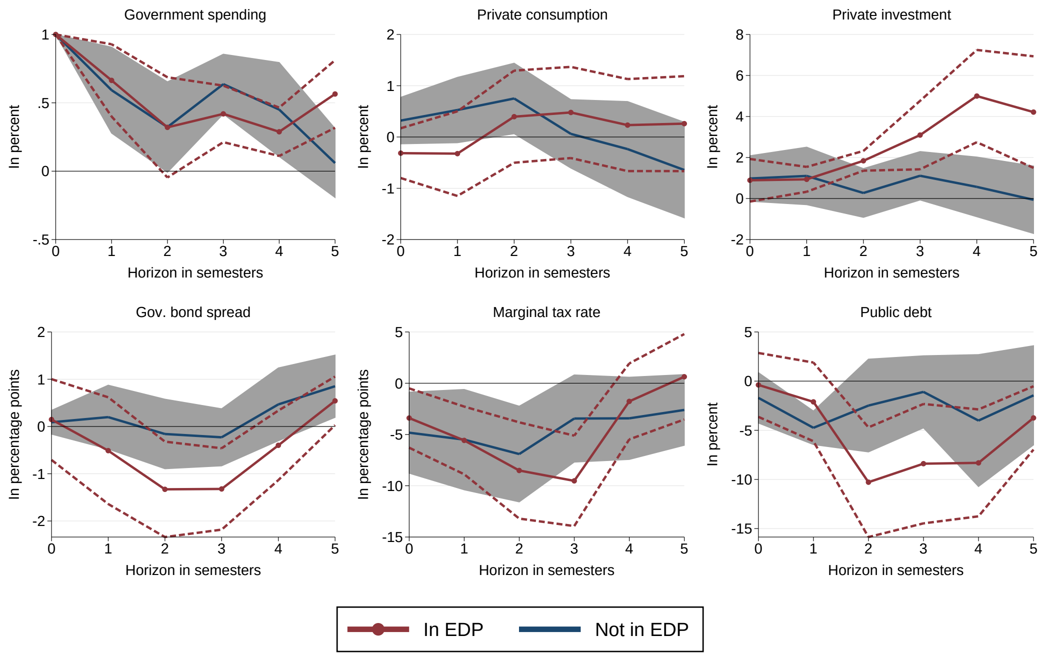

Figure 5.18. IRFs in bad times – In ECB banking crisis

Notes: Estimates in each panel are based on 122 observations. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

5.A.3 Identified Recessions

Table 5.5—Identified recessions in our sample

#	Country	Start	Length	Depth	#	Country	Start	Length	Depth
1	AUT	2001q1	2	−0.21	20	FRA	2008q2	5	−3.87
2	AUT	2012q2	4	−1.10	21	FRA	2012q4	2	−0.12
3	BEL	2001q3	2	−0.29	22	GBR	2008q2	5	−6.04
4	BEL	2008q3	3	−3.69	23	IRL	2007q2	10	−10.87
5	BEL	2012q4	2	−0.32	24	IRL	2012q3	3	−1.80
6	CZE	2008q4	3	−5.86	25	ITA	2001q2	4	−0.65
7	DEU	2001q2	4	−0.85	26	ITA	2003q1	3	−0.57
8	DEU	2002q4	10	−0.47	27	ITA	2008q2	5	−7.46
9	DEU	2008q2	4	−7.03	28	ITA	2011q3	10	−5.33
10	DEU	2012q4	2	−0.88	29	LUX	2002q3	3	−2.40
11	DNK	2001q4	3	−0.24	30	LUX	2008q1	6	−8.00
12	DNK	2006q3	4	−1.05	31	LUX	2011q2	4	−1.57
13	DNK	2008q1	6	−7.07	32	NLD	2008q3	3	−4.35
14	DNK	2011q3	6	−0.50	33	NLD	2011q2	7	−1.97
15	ESP	2008q3	4	−4.36	34	PRT	2002q2	5	−2.43
16	ESP	2011q1	11	−5.28	35	PRT	2008q2	4	−4.33
17	FIN	2008q1	6	−9.49	36	PRT	2010q4	9	−7.87
18	FIN	2012q1	4	−2.49	37	SWE	2008q1	7	−5.86
19	FIN	2013q3	4	−0.96	38	SWE	2011q4	5	−1.59

Notes: “Start” refers to the first quarter of the recession, i.e., the quarter following the peak of the business cycle. “Length” states the duration of a recession in quarters. “Depth” refers to the deviation from the pre-recession peak level of output to the trough (in %).

5.A.4 Detailed Results for Alternative States

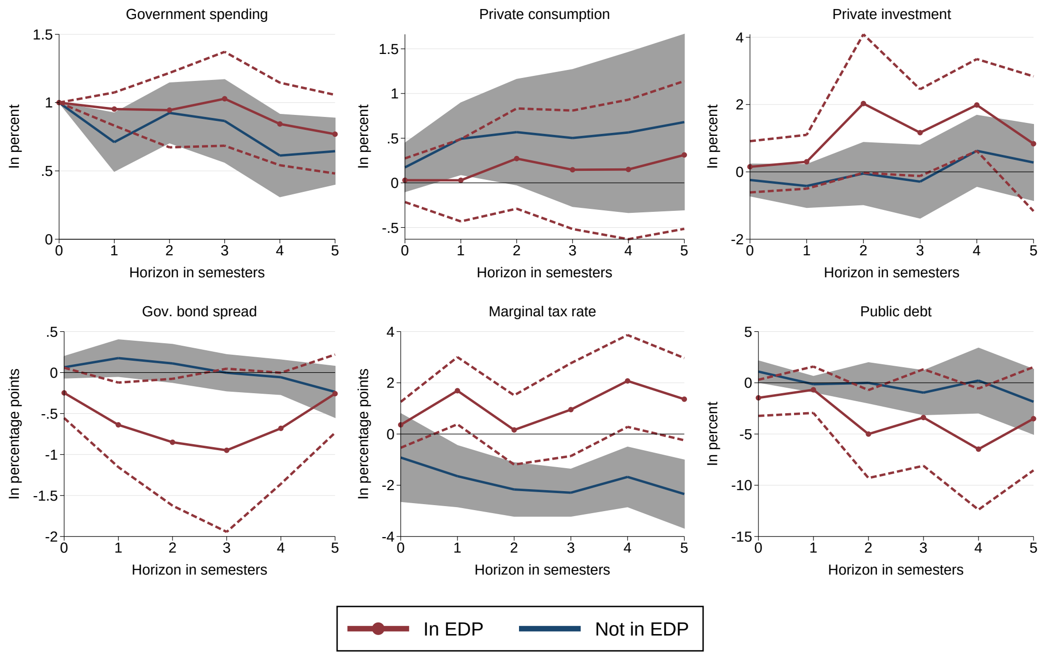

Figure 5.19. IRFs for alternative states – Maastricht criteria not fulfilled (debt or deficit)

Notes: Estimates in each panel are based on the full sample. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

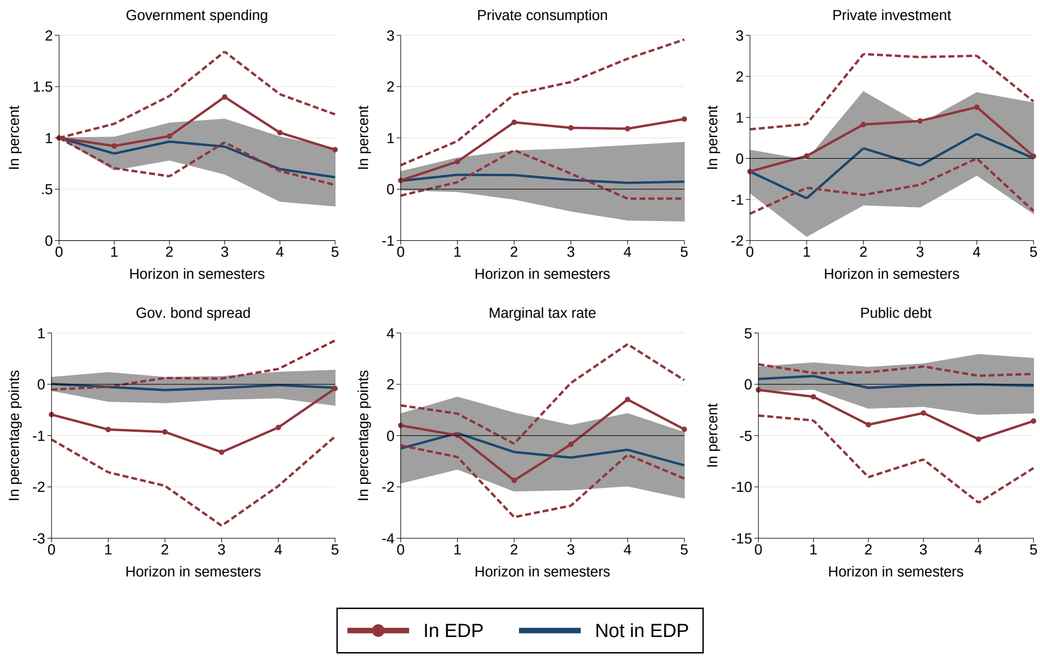

Figure 5.20. IRFs for alternative states – Maastricht criteria not fulfilled (deficit only)

Notes: Estimates in each panel are based on the full sample. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

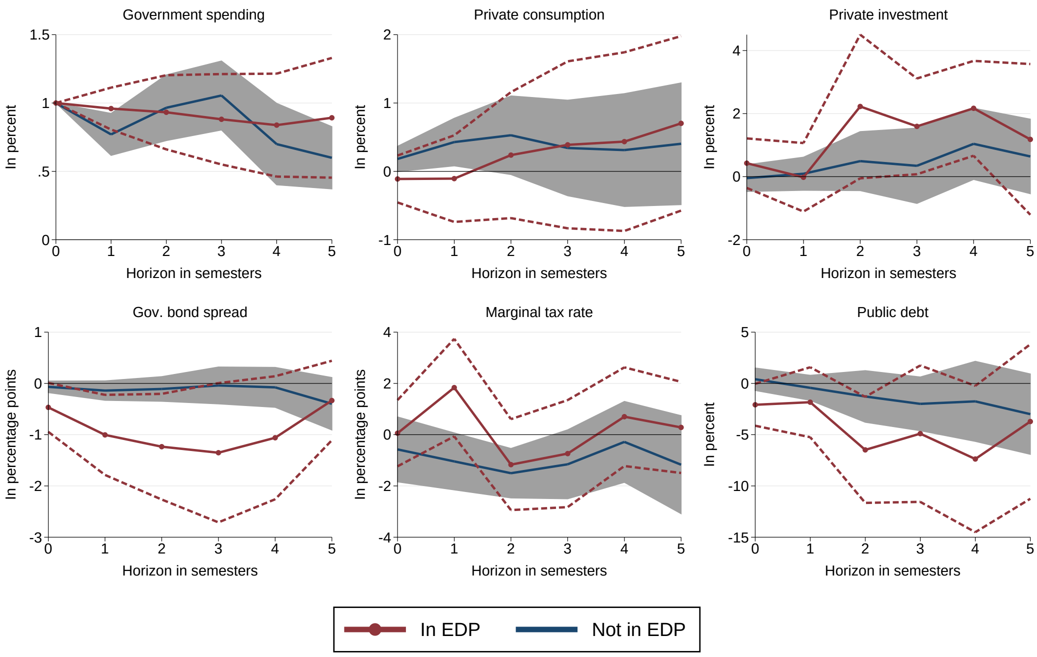

Figure 5.21. IRFs for alternative states – Maastricht criteria not fulfilled (debt only)

Notes: Estimates in each panel are based on the full sample. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

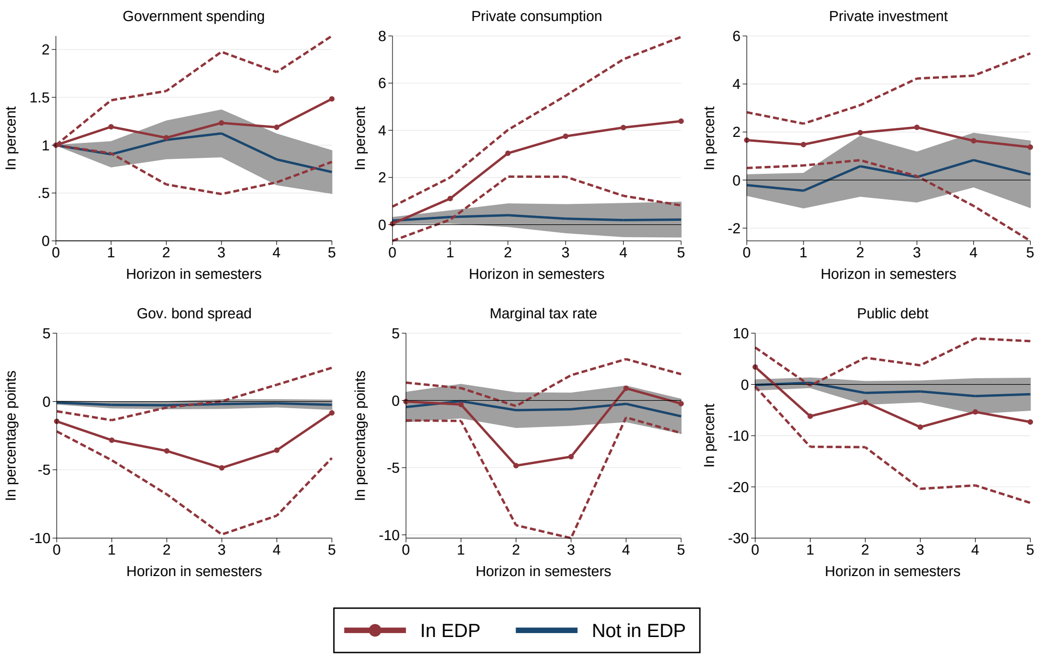

Figure 5.22. IRFs for alternative states – Maastricht criteria not fulfilled (debt and deficit)

Notes: Estimates in each panel are based on the full sample. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

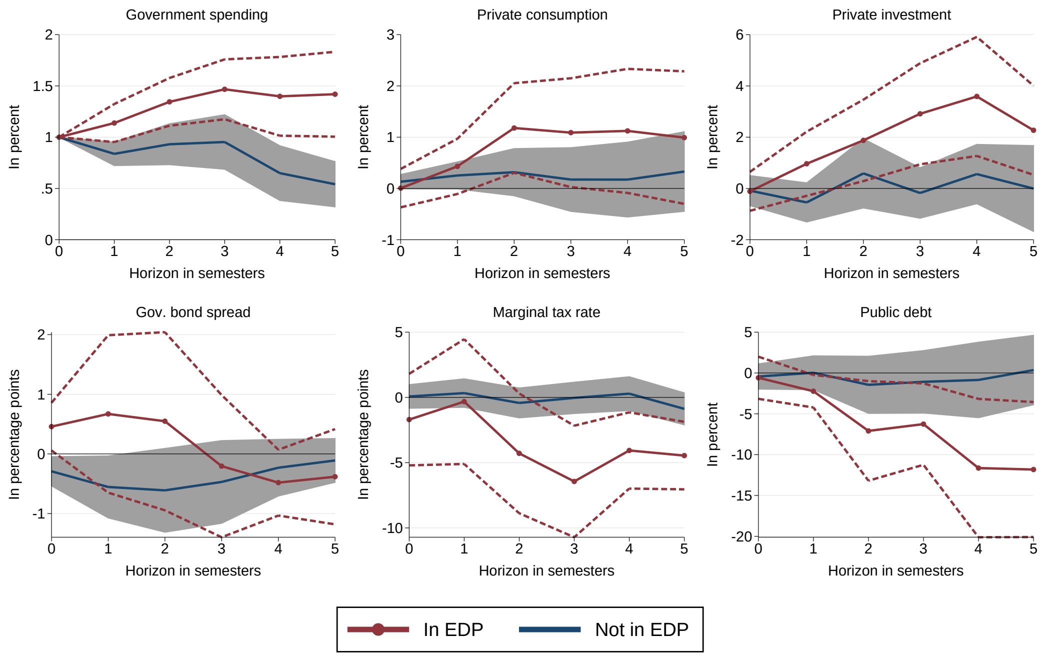

Figure 5.23. IRFs for alternative states – Recession

Notes: Estimates in each panel are based on the full sample. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

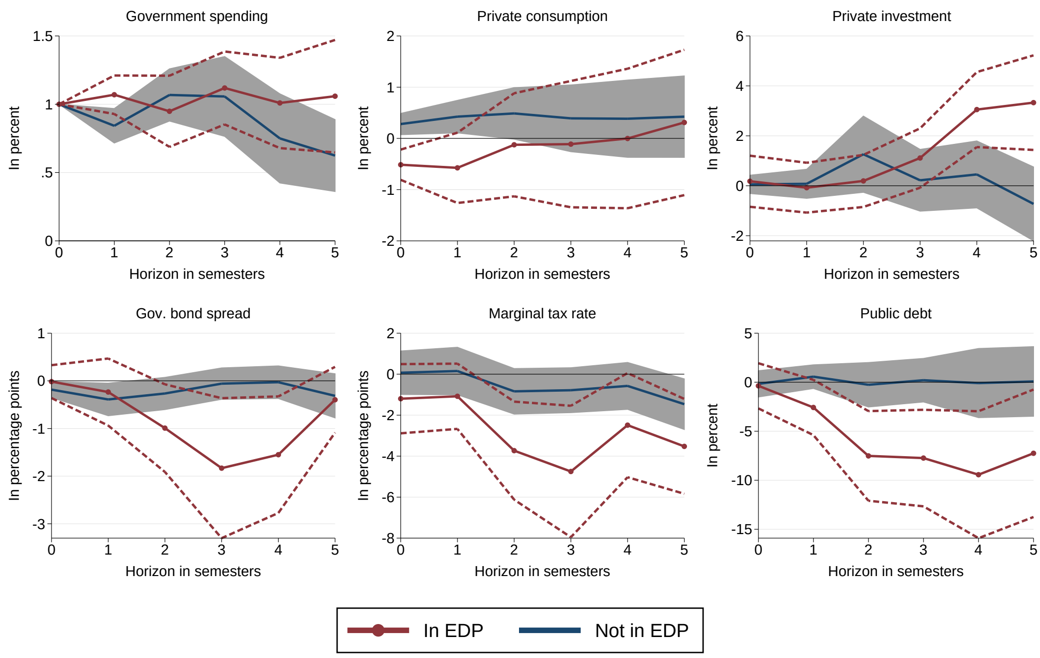

Figure 5.24. IRFs for alternative states – ECB banking crisis

Notes: Estimates in each panel are based on the full sample. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

5.A.5 Detailed Results for Real-Time Multipliers

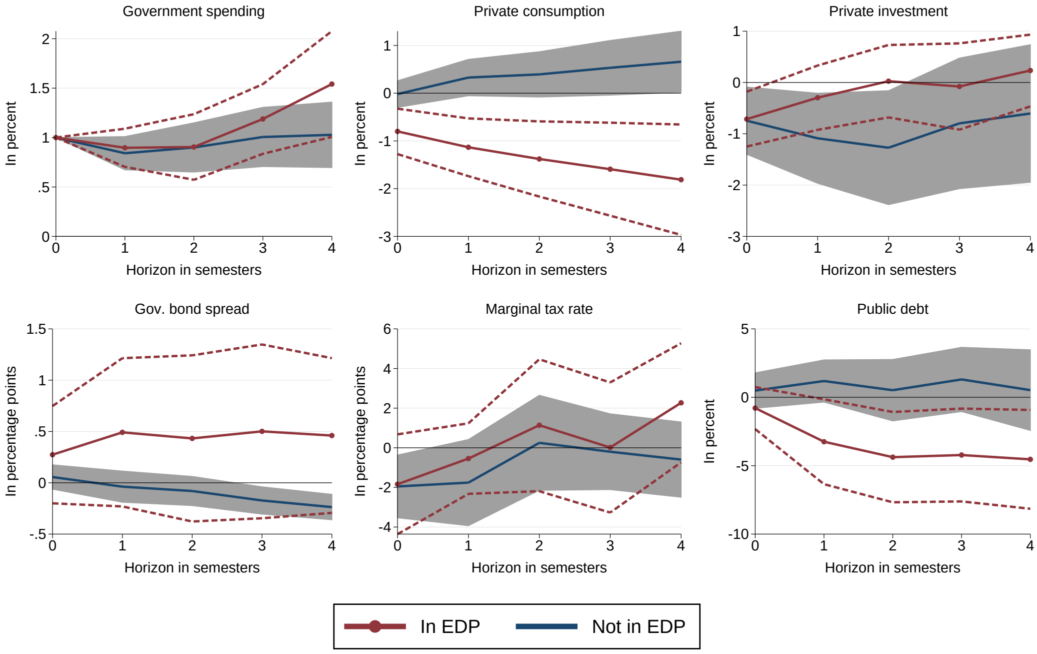

Figure 5.25. IRFs for real-time exercise $s = t + 1$

Notes: Estimates in each panel are based on 373 observations covering 15 countries. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

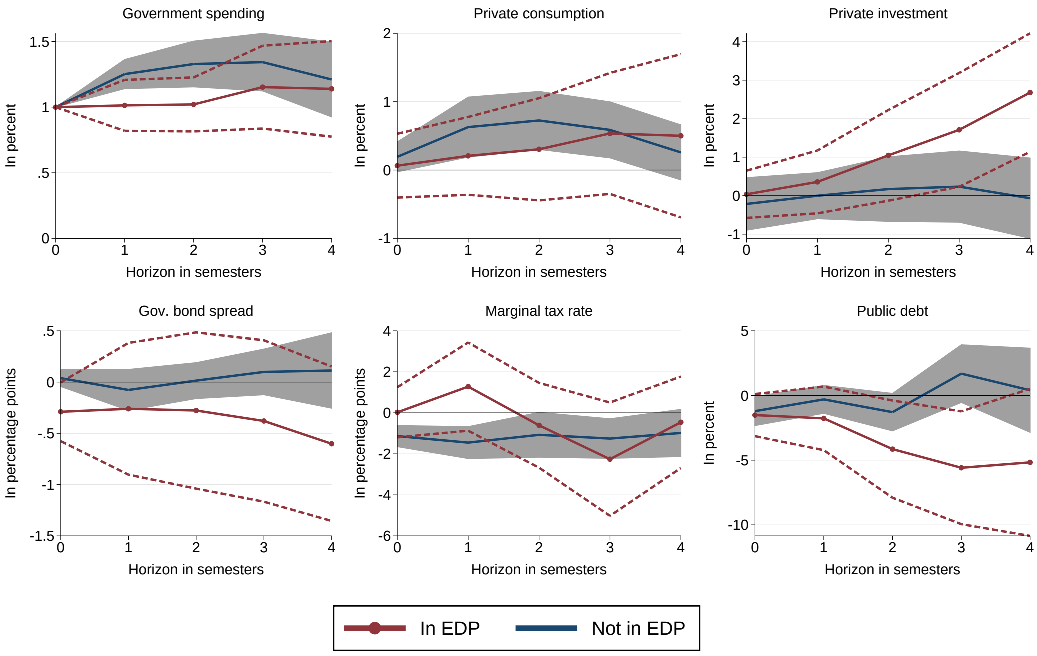

Figure 5.26. IRFs for real-time exercise $s = t + 5$

Notes: Estimates in each panel are based on 373 observations covering 15 countries. 90% confidence intervals are calculated from Driscoll-Kraay standard errors.

References

Abiad, Abdul, Petya Koeva Brooks, Irina Tytell, Daniel Leigh, and Ravi Balakrishnan (2009). *What's the Damage? Medium-term Output Dynamics after Banking Crises*. IMF Working Paper 09/245. International Monetary Fund.

Adjemian, Stéphane, Houtan Bastani, Michel Juillard, Fréderic Karamé, Ferhat Mihoubi, Willi Mutschler, Johannes Pfeifer, Marco Ratto, Normann Rion, and Sébastien Villemot (2022). *Dynare: Reference Manual Version 5*. Dynare Working Paper 72. Centre pour la recherche économique et ses applications.

Aikman, David, Andrew G. Haldane, and Benjamin D. Nelson (2015). "Curbing the Credit Cycle." *The Economic Journal* 125(585), 1072–1109.

Aizenman, Joshua, Menzie D. Chinn, and Hiro Ito (2013). "The 'Impossible Trinity' Hypothesis in an Era of Global Imbalances: Measurement and Testing." *Review of International Economics* 21(3), 447–458.

Alichi, Ali, Olivier Bizimana, Douglas Laxton, Kadir Tanyeri, Hou Wang, Jiaxiong Yao, and Fan Zhang (2017). *Multivariate filter estimation of potential output for the United States*. IMF Working Papers 17/106. International Monetary Fund.

Álvarez, Luis J. and Ana Gómez-Loscos (2018). "A menu on output gap estimation methods." *Journal of Policy Modeling* 40(4), 827–850.

An, Sungbae and Frank Schorfheide (2007). "Bayesian Analysis of DSGE Models." *Econometric Reviews* 26(2–4), 113–172.

Angerer, Jost, Martin Hradiský, Marcel Magnus, Alice Zoppé, Matteo Ciucci, Javier María Vega Bordell, and Marie Therese Bitterlich (2016). *Economic Dialogue with Ireland*. Note on 08 November. Directorate General for Internal Policies, European Parliament.

Apel, Mikael and Per Jansson (1999). "A theory-consistent system approach for estimating potential output and the NAIRU." *Economics Letters* 64(3), 271–275.

Auerbach, Alan J. and Yuriy Gorodnichenko (2012). "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4(2), 1–27.

– **(2013)**. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy after the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi, 63–98. Chicago: University of Chicago Press.

Ball, Laurence M. (2014). "Long-term damage from the Great Recession in OECD countries." *European Journal of Economics and Economic Policies* 11(2), 149–160.

Ball, Laurence M. and Sandeep Mazumder (2011). “Inflation Dynamics and the Great Recession.” *Brookings Papers on Economic Activity* 42(1), 337–405.

Banerjee, Ryan and Fabrizio Zampolli (2019). “What drives the short-run costs of fiscal consolidation? Evidence from OECD countries.” *Economic Modelling* 82, 420–436.

Barro, Robert J. and Charles J. Redlick (2011). “Macroeconomic Effects From Government Purchases and Taxes.” *Quarterly Journal of Economics* 126(1), 51–102.

Beaudry, Paul, Dana Galizia, and Franck Portier (2017). “Reconciling Hayek’s and Keynes’ Views of Recessions.” *Review of Economic Studies* 85(1), 119–156.

Beffy, Pierre-Olivier, Patrice Ollivaud, Pete Richardson, and Franck Sédillot (2006). *New OECD Methods for Supply-side and Medium-term Assessments: A Capital Services Approach*. OECD Economics Department Working Paper 482. Organisation for Economic Co-operation and Development.

Benati, Luca (2012). “Estimating the financial crisis’ impact on potential output.” *Economics Letters* 114(1), 113–119.

Benes, Jaromír, Kevin Clinton, Roberto Garcia-Saltos, Marianne Johnson, Douglas Laxton, Peter B. Manchev, and Troy Matheson (2010). *Estimating Potential Output with a Multivariate Filter*. IMF Working Papers 10/285. International Monetary Fund.

Bernardini, Marco and Gert Peersman (2018). “Private debt overhang and the government spending multiplier: Evidence for the United States.” *Journal of Applied Econometrics* 33(4), 485–508.

Blanchard, Olivier J. (2016). “The Phillips Curve: Back to the ’60s?” *American Economic Review* 106(5), 31–34.

– **(2018)**. “Should We Reject the Natural Rate Hypothesis?” *Journal of Economic Perspectives* 32(1), 97–120.

Blanchard, Olivier J., Eugenio Cerutti, and Lawrence H. Summers (2015). *Inflation and Activity – Two Explorations and their Monetary Policy Implications*. NBER Working Paper 21726. National Bureau of Economic Research.

Blanchard, Olivier J., Florence Jaumotte, and Prakash Loungani (2013). *Labor Market Policies and IMF Advice in Advanced Economies during the Great Recession*. IMF Staff Discussion Note 13/02. International Monetary Fund.

Blanchard, Olivier J. and Daniel Leigh (2013). “Growth Forecast Errors and Fiscal Multipliers.” *American Economic Review* 103(3), 117–120.

– **(2014)**. “Learning about Fiscal Multipliers from Growth Forecast Errors.” *IMF Economic Review* 62(2), 179–212.

Blanchard, Olivier J., Guido Lorenzoni, and Jean-Paul L’Huillier (2017). “Short-run effects of lower productivity growth. A twist on the secular stagnation hypothesis.” *Journal of Policy Modeling* 39(4), 639–649.

Blanchard, Olivier J. and Roberto Perotti (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.

Blanchard, Olivier J. and Lawrence H. Summers (1986). *Hysteresis in Unemployment*. NBER Working Paper 2035. National Bureau of Economic Research.

– **(1987)**. "Hysteresis in unemployment." *European Economic Review* 31(1–2), 288–295.

Boehm, Christoph E. (2020). "Government consumption and investment: Does the composition of purchases affect the multiplier?" *Journal of Monetary Economics* 115, 80–93.

Borio, Claudio (2014). "The financial cycle and macroeconomics: What have we learnt?" *Journal of Banking & Finance* 45, 182–198.

Borio, Claudio, Piti Disyatat, and Mikael Juselius (2014). *A parsimonious approach to incorporating economic information in measures of potential output*. BIS Working Paper 442. Bank for International Settlements.

– **(2017)**. "Rethinking potential output: embedding information about the financial cycle." *Oxford Economic Papers* 69(3), 655–677.

Brooks, Stephen P. and Andrew Gelman (1998). "General Methods for Monitoring Convergence of Iterative Simulations." *Journal of Computational and Graphical Statistics* 7(4), 434–455.

Bry, Gerhard and Charlotte Boschan (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. NBER Books. National Bureau of Economic Research.

CBO (2001). *CBO's Method for Estimating Potential Output: An Update*. Report.

– **(2014)**. *Revisions to CBO's Projection of Potential Output Since 2007*. Report 4688.

Camba-Mendez, Gonzalo and Diego Rodriguez-Palenzuela (2003). "Assessment criteria for output gap estimates." *Economic Modelling* 20(3), 529–562.

Cerra, Valerie and Sweta C. Saxena (2008). "Growth Dynamics: The Myth of Economic Recovery." *American Economic Review* 98(1), 439–457.

Chinn, Menzie D. and Hiro Ito (2006). "What matters for financial development? Capital controls, institutions, and interactions." *Journal of Development Economics* 81(1), 163–192.

Claessens, Stijn, M. Ayhan Kose, and Marco E. Terrones (2012). "How do business and financial cycles interact?" *Journal of International Economics* 87(1), 178–190.

Clarida, Richard, Jordi Galí, and Mark Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115(1), 147–180.

Clark, Peter K. (1987). "The Cyclical Component of U. S. Economic Activity." *Quarterly Journal of Economics* 102(4), 797–814.

– **(1989)**. "Trend reversion in real output and unemployment." *Journal of Econometrics* 40(1), 15–32.

Coibion, Olivier, Yuriy Gorodnichenko, and Mauricio Ulate (2018). "The Cyclical Sensitivity in Estimates of Potential Output." *Brookings Papers on Economic Activity* 49(2), 343–441.

Corsetti, Giancarlo, André Meier, and Gernot J. Müller (2012a). "Fiscal stimulus with spending reversals." *Review of Economics and Statistics* 94(4), 878–895.

– **(2012b)**. "What determines government spending multipliers?" *Economic Policy* 27(72), 521–565.

Corsetti, Giancarlo, Paolo Pesenti, and Nouriel Roubini (1999). "What caused the Asian currency and financial crisis?" *Japan and the World Economy* 11(3), 305–373.

Croushore, Dean and Tom Stark (2001). "A real-time data set for macroeconomists." *Journal of Econometrics* 105(1), 111–130.

D'Auria, Francesca, Cécile Denis, Karel Havik, Kieran McMorrow, Christophe Planas, Rafal Raciborski, Werner Röger, and Alessandro Rossi (2010). *The production function methodology for calculating potential growth rates and output gaps*. European Economy Economic Paper 420. European Commission.

Darvas, Zsolt and András Simon (2015). *Filling the gap: Open economy considerations for more reliable potential output*. Bruegel Working Paper 15/11. Bruegel.

De Jong, Jasper F. M. and Niels D. Gilbert (2020). "Fiscal discipline in EMU? Testing the effectiveness of the Excessive Deficit Procedure." *European Journal of Political Economy* 61, 101822.

Denis, Cécile, Daniel Grenouilleau, Kieran McMorrow, and Werner Röger (2006). *Calculating potential growth rates and output gaps – A revised production function approach*. European Economy Economic Paper 247. European Commission.

Diebold, Francis X. and Roberto S. Mariano (1995). "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13(3), 253–263.

Dovern, Jonas, Ulrich Fritsche, Prakash Loungani, and Natalia Tamirisa (2015). "Information rigidities: Comparing average and individual forecasts for a large international panel." *International Journal of Forecasting* 31(1), 144–154.

Dovern, Jonas and Johannes Weisser (2011). "Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7." *International Journal of Forecasting* 27(2), 452–465.

Dovern, Jonas and Christopher Zuber (2020a). "Recessions and Potential Output: Disentangling Measurement Errors, Supply Shocks, and Hysteresis Effects." *The Scandinavian Journal of Economics* 122(4), 1431–1466.

– **(2020b)**. "How economic crises damage potential output – Evidence from the Great Recession." *Journal of Macroeconomics* 65, 103239.

Draghi, Mario (2017). *Accompanying the economic recovery*. Introductory speech by the president of the ECB at the ECB Forum on Central Banking, Sintra, June 27. URL: `https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp170627.en.html` (visited on July 6, 2020).

– **(2019)**. *Stabilisation policies in a monetary union*. Speech by the president of the ECB at the Academy of Athens, Athens, October 1. URL: `https://www.ecb.europa.eu/press/key/date/2019/html/ecb.sp191001_1~5d7713fcd1.en.html` (visited on July 6, 2020).

Driscoll, John C. and Aart C. Kraay (1998). “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data.” *Review of Economics and Statistics* 80(4), 549–560.

Dumitrescu, Elena-Ivona and Christophe Hurlin (2012). “Testing for Granger non-causality in heterogeneous panels.” *Economic Modelling* 29(4), 1450–1460.

Edge, Rochelle M. and Jeremy B. Rudd (2016). “Real-Time Properties of the Federal Reserve’s Output Gap.” *Review of Economics and Statistics* 98(4), 785–791.

Enders, Zeno, Michael Kleemann, and Gernot J. Müller (2018). *Growth expectations, undue optimism, and short-run fluctuations*. Working Paper.

Erceg, Christopher J. and Andrew T. Levin (2014). “Labor Force Participation and Monetary Policy in the Wake of the Great Recession.” *Journal of Money, Credit and Banking* 46(2), 3–49.

European Commission (2008). *Economic Forecast, Spring 2008*. European Commission, Luxembourg.

European Union (2008). *Treaty on the Functioning of the European Union*. Legislative text. URL: `https://eur-lex.europa.eu/eli/treaty/tfeu_2008/art_126/oj` (visited on March 31, 2021).

Fatás, Antonio and Lawrence H. Summers (2018). “The permanent effects of fiscal consolidations.” *Journal of International Economics* 112, 238–250.

Fernald, John G. (2015). “Productivity and Potential Output before, during, and after the Great Recession.” *NBER Macroeconomics Annual* 29, 1–51.

Furceri, Davide and Annabelle Mourougane (2012). “The effect of financial crises on potential output: New empirical evidence from OECD countries.” *Journal of Macroeconomics* 34(3), 822–832.

Gadea Rivas, Maria Dolores and Gabriel Perez-Quiros (2015). “The Failure to Predict the Great Recession—A View through the Role of Credit.” *Journal of the European Economic Association* 13(3), 534–559.

Galí, Jordi (2016). *Insider-Outsider Labor Markets, Hysteresis and Monetary Policy*. Economics Working Paper 1506. Department of Economics and Business, Universitat Pompeu Fabra.

Garratt, Anthony, Kevin Lee, Emi Mise, and Kalvinder Shields (2008). “Real-Time Representations of the Output Gap.” *Review of Economics and Statistics* 90(4), 792–804.

Gerlach, Stefan and Frank Smets (1999). "Output gaps and monetary policy in the EMU area." *European Economic Review* 43(4), 801–812.

Gneiting, Tilmann and Adrian E. Raftery (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102(477), 359–378.

González-Astudillo, Manuel (2019a). "An output gap measure for the euro area: Exploiting country-level and cross-sectional data heterogeneity." *European Economic Review* 120, 103301.

– **(2019b)**. "Estimating the U.S. output gap with state-level data." *Journal of Applied Econometrics* 34(5), 795–810.

Gordon, Robert and Robert Krenn (2010). *The End of the Great Depression 1939-41: Policy Contributions and Fiscal Multipliers*. NBER Working Paper 16380. National Bureau of Economic Research.

Górnicka, Lucyna, Christophe Kamps, Gerrit Koester, and Nadine Leiner-Killinger (2020). "Learning about fiscal multipliers during the European sovereign debt crisis: evidence from a quasi-natural experiment." *Economic Policy* 35(101), 5–40.

Guisinger, Amy Y., Michael T. Owyang, and Hannah G. Shell (2018). "Comparing Measures of Potential Output." *Federal Reserve Bank of St. Louis Review* 100(4), 297–316.

Hagen, Jürgen von and Barry Eichengreen (1996). "Federalism, Fiscal Restraints, and European Monetary Union." *American Economic Review* 86(2), 134–138.

Hall, Robert E. (2009). "By How Much Does GDP Rise If the Government Buys More Output?" *Brookings Papers on Economic Activity* 40(2), 183–249.

– **(2014)**. "Quantifying the Lasting Harm to the U.S. Economy from the Financial Crisis." *NBER Macroeconomics Annual* 29, 71–128.

Haltmaier, Jane (2012). *Do Recessions Affect Potential Output?* International Finance Discussion Paper 1066. Board of Governors of the Federal Reserve System.

Hamilton, James D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

– **(2018)**. "Why You Should Never Use the Hodrick-Prescott Filter." *Review of Economics and Statistics* 100(5), 831–843.

Hansen, Nikolaus and Stefan Kern (2004). "Evaluating the CMA Evolution Strategy on Multimodal Test Functions." In *Parallel Problem Solving from Nature - PPSN VIII*. Edited by Xin Yao, Edmund K. Burke, José A. Lozano, Jim Smith, Juan Julián Merelo-Guervós, John A. Bullinaria, Jonathan E. Rowe, Peter Tiňo, Ata Kabán, and Hans-Paul Schwefel. Berlin, Heidelberg: Springer, 282–291.

Harding, Don and Adrian Pagan (2002). "Dissecting the cycle: a methodological investigation." *Journal of Monetary Economics* 49(2), 365–381.

Harvey, Andrew C. (1985). "Trends and Cycles in Macroeconomic Time Series." *Journal of Business & Economic Statistics* 3(3), 216–227.

Harvey, Andrew C. and Albert Jaeger (1993). "Detrending, Stylized Facts and the Business Cycle." *Journal of Applied Econometrics* 8(3), 231–247.

Havik, Karel, Kieran McMorrow, Fabrice Orlandi, Christophe Planas, Rafal Raciborski, Werner Röger, Alessandro Rossi, Anna Thum-Thysen, and Valerie Vandermeulen (2014). *The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps*. European Economy Economic Paper 535. European Commission.

Heimberger, Philipp, Jakob Kapeller, and Bernhard Schütz (2017). "The NAIRU determinants: What's structural about unemployment in Europe?" *Journal of Policy Modeling* 39(5), 883–908.

Hodrick, Robert J. (2020). *An Exploration of Trend-Cycle Decomposition Methodologies in Simulated Data*. NBER Working Paper 26750. National Bureau of Economic Research.

Hosseinkouchack, Mehdi and Maik H. Wolters (2013). "Do large recessions reduce output permanently?" *Economics Letters* 121(3), 516–519.

Huidrom, Raju, M. Ayhan Kose, Jamus J. Lim, and Franziska L. Ohnsorge (2020). "Why do fiscal multipliers depend on fiscal Positions?" *Journal of Monetary Economics* 114, 109–125.

Ilzetzki, Ethan, Enrique G. Mendoza, and Carlos A. Végh (2013). "How big (small?) are fiscal multipliers?" *Journal of Monetary Economics* 60(2), 239–254.

Jacobs, Jan P.A.M. and Simon van Norden (2016). "Why are initial estimates of productivity growth so unreliable?" *Journal of Macroeconomics* 47(Part B), 200–213.

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

Jordan, Alexander, Fabian Krüger, and Sebastian Lerch (2019). "Evaluating Probabilistic Forecasts with scoringRules." *Journal of Statistical Software* 90, 1–37.

Kass, Robert E. and Adrian E. Raftery (1995). "Bayes Factors." *Journal of the American Statistical Association* 90(430), 773–795.

Kiley, Michael T. (2013). "Output gaps." *Journal of Macroeconomics* 37, 1–18.

Koopman, Siem J. and James Durbin (2003). "Filtering and smoothing of state vector for diffuse state-space models." *Journal of Time Series Analysis* 24(1), 85–98.

Krugman, Paul (2011). "Can Europe Be Saved?" *New York Times Magazine*, January 12. URL: `https://www.nytimes.com/2011/01/16/magazine/16Europe-t.html` (visited on July 6, 2020).

Kuttner, Kenneth N. (1994). "Estimating Potential Output as a Latent Variable." *Journal of Business & Economic Statistics* 12(3), 361–368.

Laeven, Luc and Fabián Valencia (2013). "Systemic Banking Crises Database." *IMF Economic Review* 61(2), 225–270.

Lendvai, Julia, Salto Matteo, and Anna Thum-Thysen (2015). *Structural unemployment vs. NAWRU: Implications for the assessment of the cyclical position and the fiscal stance*. European Economy Economic Paper 552. European Commission.

Lindbeck, Assar and Dennis J. Snower (1986). "Wage Setting, Unemployment, and Insider-Outsider Relations." *American Economic Review* 76(2), 235–39.

Lo Duca, Marco, Anne Koban, Marisa Basten, Elias Bengtsson, Benjamin Klaus, Piotr Kusmierczyk, Jan Hannes Lang, Carsten Detken, and Tuomas A. Peltonen (2017). *A new database for financial crises in European countries*. ECB Occasional Paper 194. European Central Bank.

Marcellino, Massimiliano and Alberto Musso (2011). "The reliability of real-time estimates of the euro area output gap." *Economic Modelling* 28(4), 1842–1856.

Martin, Robert F., Teyanna Munyan, and Beth Anne Wilson (2015). *Potential Output and Recessions: Are We Fooling Ourselves?* International Finance Discussion Paper 1145. Board of Governors of the Federal Reserve System.

Melolinna, Marko and Máté Tóth (2019). "Output gaps, inflation and financial cycles in the UK." *Empirical Economics* 56(3), 1039–1070.

Mineshima, Aiko, Marcos Poplawski-Ribeiro, and Anke Weber (2014). "Size of Fiscal Multipliers." In *Post-crisis Fiscal Policy*, edited by Carlo Cottarelli, Philip Gerson, and Abdelhak Senhadji, 315–372. Cambridge, MA: MIT Press.

Montiel Olea, José Luis and Carolin Pflueger (2013). "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31(3), 358–369.

Morley, James and Jeremy Piger (2012). "The Asymmetric Business Cycle." *Review of Economics and Statistics* 94(1), 208–221.

Mountford, Andrew and Harald Uhlig (2009). "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics* 24(6), 960–992.

Nickel, Christiane and Andreas Tudyka (2014). "Fiscal Stimulus in Times of High Debt: Reconsidering Multipliers and Twin Deficits." *Journal of Money, Credit and Banking* 46(7), 1313–1344.

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

OECD (1998). *OECD Economic Outlook, Volume 1998 Issue 1*. OECD Publishing, Paris.

– **(2001)**. *OECD Economic Outlook, Volume 2001 Issue 2*. OECD Publishing, Paris.

– **(2016)**. *OECD Economic Outlook, Volume 2016 Issue 2*. OECD Publishing, Paris.

Okun, Arthur M. (1962). *Potential GNP - Its Measurement and Significance*. Reprint as Cowles Foundation Paper 190. Yale University.

Orphanides, Athanasios (2003). "Monetary policy evaluation with noisy information." *Journal of Monetary Economics* 50(3), 605–631.

Orphanides, Athanasios, Richard D. Porter, David Reifschneider, Robert Tetlow, and Frederico Finan (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.

Orphanides, Athanasios and Simon van Norden (2002). "The Unreliability of Output-Gap Estimates in Real Time." *Review of Economics and Statistics* 84(4), 569–583.

Oulton, Nicholas and María Sebastiá-Barriel (2017). "Effects of Financial Crises on Productivity, Capital and Employment." *Review of Income and Wealth* 63(s1), S90–S112.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy (2013). "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data." *American Economic Review* 103(3), 129–134.

Papell, David H. and Ruxandra Prodan (2012). "The Statistical Behavior of GDP after Financial Crises and Severe Recessions." *The B.E. Journal of Macroeconomics* 12(3), 1–31.

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

Reinhart, Carmen M. and Kenneth S. Rogoff (2009). "The Aftermath of Financial Crises." *American Economic Review* 99(2), 466–472.

– **(2014)**. "Recovery from Financial Crises: Evidence from 100 Episodes." *American Economic Review* 104(5), 50–55.

Romer, Christina D. and David H. Romer (2018). "Phillips Lecture – Why Some Times Are Different: Macroeconomic Policy and the Aftermath of Financial Crises." *Economica* 85(337), 1–40.

Schularick, Moritz and Alan M. Taylor (2012). "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008." *American Economic Review* 102(2), 1029–1061.

Stadler, George W. (1986). "Real versus monetary business cycle theory and the statistical characteristics of output fluctuations." *Economics Letters* 22(1), 51–54.

– **(1990)**. "Business Cycle Models with Endogenous Technology." *American Economic Review* 80(4), 763–778.

Stock, James H. and Mark W. Watson (1999). "Forecasting inflation." *Journal of Monetary Economics* 44(2), 293–335.

Watson, Mark W. (1986). "Univariate detrending methods with stochastic trends." *Journal of Monetary Economics* 18(1), 49–75.

Yellen, Janet L. (2017). *The Economic Outlook and the Conduct of Monetary Policy*. Speech by the chair of the Federal Reserve at the Stanford Institute for Economic Policy Research, Stanford, California, January 19. URL: `https://www.federalreserve.gov/newsevents/speech/yellen20170119a.htm` (visited on July 6, 2020).