CESifo WORKING PAPERS

8911
2021

February 2021

Survey-Based Structural Budget Balances

Marcell Göttert, Timo Wollmershäuser

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8911*

Survey-Based Structural Budget Balances

Abstract

The budget dispute between Italy and the European Commission in 2018 gave new impetus for the debate about the reliability of output gap estimation methods and their use for calculating structural budget balances. In this paper we review the main properties of the mainstream approaches and compare their performance with structural budget balances, whose calculation is based on a business survey. Our main result is that while the survey-based measure is highly correlated with the existing structural budget balances which are calculated based on some estimates of the output gap, it is significantly less revised over time and almost unbiased. Moreover, the survey-based measure could be easily implemented into the existing EU fiscal rules without any major changes.

JEL Codes: E320, E620, H620, H680.

Keywords: fiscal rules, cyclical adjustment, output gaps, real time data.

*Marcell Göttert**
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5
Germany – 81679 Munich
goettert@ifo.de

Timo Wollmershäuser
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5
Germany – 81679 Munich
wollmershaeuser@ifo.de

*corresponding author

19th February 2021
We thank the participants of the ifo Macro Seminar, the CIRET Workshop 2019, the ifo Christmas Conference 2019, the NOeG Winter Workshop 2019 and the NOeG Annual Meeting 2020 for valuable comments.

1 Introduction

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

In this paper we investigate the structural budget balances published by the EC, the Organisation for Economic Co-operation and Development (OECD) and the International Monetary Fund (IMF). The three institutions give economic policy recommendations to their member countries and analyze their economic situation. These recommendations and analyses also refer to fiscal policy and include fiscal surveillance. We show that the structural budget balances of theses institutions, which are all estimated on the basis of a production function, are indeed systematically revised downward after their initial release. The source of this bias is the underlying output gap estimation which suffers from decomposing the level of GDP into a stable trend and cycle. Thus, critics are right when they argue that the production function-based approach of estimating structural budget balances is an inappropriate tool for fiscal surveillance.

Therefore, we propose a survey-based approach that avoids the shortcomings of the production function-based approach. Our measure is based on the degree of capacity utilization of firms, which is calculated from a representative business survey. It is available without publication lag in the middle of the current quarter and is highly correlated with the existing estimates for the output gaps. We show that in comparison with the production function-based approach structural budget balances calculated using survey-based capacity utilization are significantly less biased and hardly revised. This high reliability, precision and early availability make our new estimation method a perfect tool for fiscal surveillance and the EU fiscal framework. It could be easily implemented into the existing EU fiscal rules without any major changes.

There are a number of papers on the evaluation and the construction of output gaps. Some of them focus on the use of capacity utilization in the manufacturing sector for improving output gap estimates, but most of them are silent regarding the consequences of output gap revisions on structural budget balances. Kempkes (2014) evaluates the output gaps published by the EC, the OECD and the IMF. His analysis concentrates on the revisions of output gap forecasts, but ignores revisions of ex post output gaps. Moreover, he argues that the biased output gap forecasts can lead to significant additional public debt. Marcellino and Musso (2011) show that the results of most of the output gap estimation methods for the euro area are very uncertain and unstable over time. They recommend the use of capacity utilization to determine the output gap and attest it both, high stability and low uncertainty. Orphanides *et al.* (2000) show the high correlation between output gap estimates and capacity utilization. Nyman (2010), Hulej and

Grabek (2015), Szörfi (2015) and Silva *et al.* (2016) include capacity utilization or other survey data as cyclical information into their empirical models for decomposing the level of GDP into trend and cycle. Their survey-based measures have smaller revisions of the cyclical component than the traditional methods. In this paper we follow the recommendation of Marcellino and Musso (2011) and directly use capacity utilization as a measure for the output gap. We construct an aggregate capacity utilization by combining survey data of the manufacturing and the service sector and apply it to the estimation of ex post structural budget balances.

The remainder of this paper is organized as follows. Section 2 gives a short introduction to the concept of structural budget balances. Sections 3 and 4 evaluate the mainstream methods of estimating structural budget balances. Section 5 introduces a survey-based method and compares it to the mainstream approach. Section 6 concludes.

2 The Structural Budget Balance

Structural budget balances are used to evaluate the budget balance without the distorting influences of the business cycle and temporary one-offs. They are mostly used for fiscal surveillance and of utmost importance for the EU fiscal rules. One of the main parts of the fiscal framework in the EU is laid out by the Stability and Growth Pact (SGP), which consists of two branches: a preventive and a corrective arm. While the preventive arm ensures fiscal stability by preventing an excessive deficit and excessive debt, the corrective arm helps a member state out of an excessive deficit or excessive debt. In the preventive arm, some of the spending goals are formulated in terms of the structural budget balance and evaluated ex post. In the corrective arm, the path out of an excessive deficit or debt is a combination of an expenditure path and a structural budget balance path, which is also evaluated ex post. Moreover, the requirements for the expenditure path and consequently the reduction of the debt level depend on the output gap (ECOFIN, 2017; EC, 2018). However, the fiscal policy of the EU member states is not only regulated by the SGP but also by the fiscal compact. There, it is set out the signatory states may not have a structural deficit larger than 0.5% of gross domestic product (EU, 2012).[1]

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

$$X = B - \epsilon_{B,Y} Y - T. \tag{1}$$

In the approach used by the EC, the OECD and the IMF the business cycle Y is measured by the output gap (GAP), which is defined as the percentage deviation of real gross domestic product (GDP) from potential output (POT):

$$GAP = \frac{GDP - POT}{POT} \cdot 100\%. \tag{2}$$

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

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

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

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

3 Revisions of the Structural Budget Balance: Production Function Approach

3.1 Data and Revision Size

For the evaluation of the production function approach we use data releases of the EC, the OECD and the IMF for Germany (DE), France (FR), United Kingdom (UK), Italy (IT), Spain (ES) and Austria (AT). Each institution has two data releases per year, one in spring and one in autumn. For each country and each institution we use the vintages ranging from spring 2003 until spring 2018. The real-time vintages of the EC are downloaded from the EC's Communication and Information Resource Centre for Administrations, Businesses and Citizens (CIRCABC). The real time data of the OECD and the IMF has been extracted from their biannual analyses, the OECD Economic Outlook and the IMF World Economic Outlook.[2] For every institution, we use their country-specific semi-elasticity to calculate the structural budget balance. The semi-elasticities are available in Mourre *et al.* (2019) for the EC, Girouard and André (2005) for the OECD and IMF (1993) for the IMF. [3]

We focus our analysis on ex post revisions of the structural budget balance. We denote by $X_{k,l,t+i}$ the ith ex post estimation of the structural budget balance of institution k for country l and year t. The first ex post estimation for year t is released in spring of the year following t (i.e. i=1), the second in autumn of the year following t (i.e. i=2), and so on. We assume that ex post revisions of the structural budget balance are only driven by revisions of the cyclical component and hence the estimation of the output gap. Thus, ex post data revisions of the budget balance B or the one-offs T, or revisions of the semi-elasticity $\epsilon_{B,GAP}$ are excluded from the analysis.[4] For $j > i$ the ex post revisions of the structural budget balance are then given by

$$X_{k,l,t+i} - X_{k,l,t+j} = \epsilon_{B,GAP,k,l}(GAP_{k,l,t+i} - GAP_{k,l,t+j}). \quad (3)$$

Figure 1 shows the size of the revisions, which is calculated for each country as the mean over all years t ranging from 2002 to 2017 and all institutions k of the absolute ex post revision of the structural budget balances between two subsequent vintages $t+i$ and $t+i+1$. The size of the revisions is sizable and persistent. It is 0.28 percentage points of potential output for the first revision for Spain, 0.20 for France, 0.19 for the UK, 0.17 for Italy, 0.14 for Germany and 0.13 for Austria. The revision size declines only slowly for later revisions. For the eighth revision (autumn release four years after t compared to the spring release four years after t) it is still between 0.06 and 0.11 percentage points of potential output. The largest revisions are found for Spain. To some extent these revisions are caused by revisions of ex post GDP data, but the largest share is due to forecast errors and forecast revisions that are used to mitigate the endpoint problem.[5]

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

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

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

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

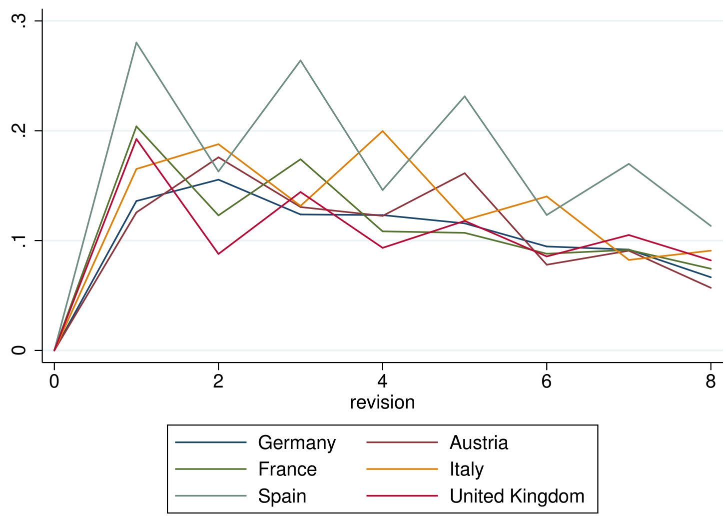

Figure 1: Size of the ex post revisions of the structural budget balance between two subsequent vintages using the production function approach (in percentage points of potential output)

3.2 Biasedness

Large revisions of ex post structural budget balances make the application and implementation of EU fiscal rules difficult and the recommendations based on them uncertain and contestable. However, the problem becomes even bigger, if the revisions not only turn out to be large, but systematic and biased. To investigate this we follow Kempkes (2014) and apply an empirical model that was developed by Holden and Peel (1990):

$$X_{l,t+1} - X_{l,t+j} = \alpha_{l,j} + u_{l,j}, \tag{4}$$

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

In an ideal world, the estimation method for the structural budget balance should on average correctly assesses the cyclical component of the budget balance already with the initial estimation. In such a case α in equation (4) would be equal zero. If, however, α turns out to be significantly different from zero, the estimation method for the structural budget balance is biased. If $\alpha < 0$, the structural budget balance X initially turns out to be overestimated and is revised downward in later estimations. These revisions can be traced back to a systematic underestimation of the output gap and consequently the cyclical component of the budget balance. If $\alpha > 0$, the opposite is the case. The structural budget balance X would be systematically revised upward. The output gap would initially be overestimated and the structural budget balance underestimated. So, the structural budget balance in the initial estimation would be systematically lower than in the later ones.

The estimation results for equation (4) are shown in Figure 2. The red line plots the estimates of α against the number of revisions $j - 1$. The shaded area around the red line is the 95%

confidence interval. For all countries the estimated structural budget balances are systematically revised downward, implying that the initial ex post release is too optimistic. Hence, in $t+1$ the fiscal position for year t looks better than a few years later. This overestimation is the result of an underestimation of the output gap and hence of the cyclical component of the budget balance. As the output gap for a certain year t is systematically larger in $t+j$ than in $t+1$, the structural budget balance for year t is smaller in $t+j$ than in $t+1$.

For Italy and France Figure 2 shows that the structural budget balance gets systematically revised downward by 0.8 percentage points of potential output after 8 revisions, i.e. after 4 years. In the United Kingdom, Germany, Austria and Spain the downward revision is somewhat less pronounced with 0.3 to 0.5 percentage points.[6] Apart from Spain all downward revision after 4 years are statistically significant at a 5% level, implying that the method for estimating the structural budget balance is biased. For Italy, France and the United Kingdom this bias is already significant after the first revision, for Germany after four revisions and for Austria after five revisions. The large confidence bands for the Spanish bias reflects the huge revisions of the structural budget balance estimations for Spain, that have already been detected in the previous Section.

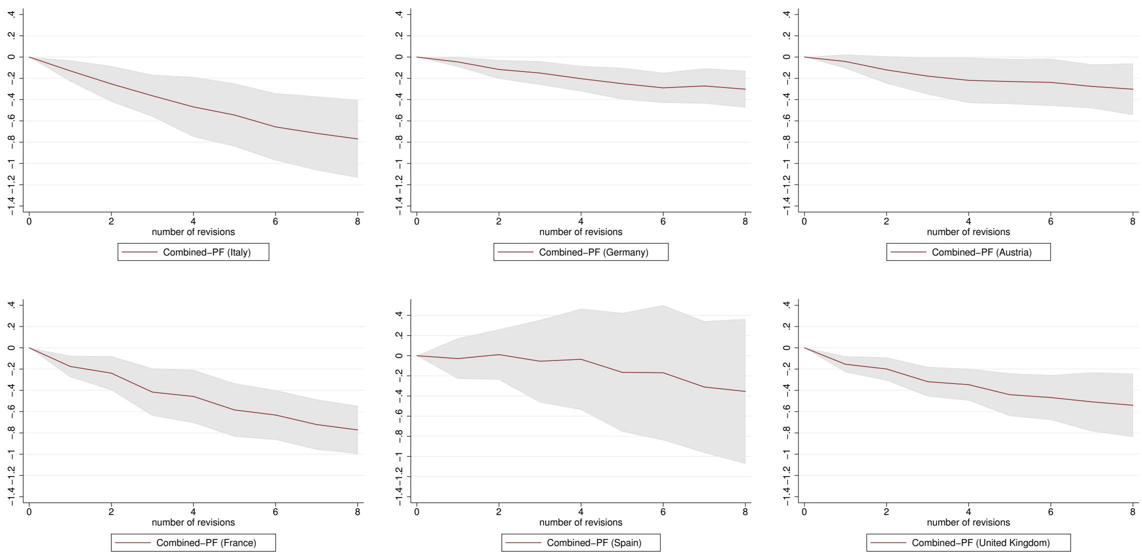

Figure 2: Bias of the ex post revisions of the structural budget (including 95% confidence interval) with respect to the initial estimate using the production function approach (in % of potential output)

Considering the importance of structural budget balances for the general fiscal surveillance and especially for the EU fiscal rules, even a small bias can be seen as highly problematic. Thus, a structural budget balance for year t that has been in line with the EU fiscal rules in $t+1$, might not be in line any more some time later in $t+j$ with $j>1$. This finding of an overestimation of the structural budget balance and its large revisions is in line with the findings of Kempkes

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

(2014), Marcellino and Musso (2011) and others, who focused their analyses on output gaps and showed that these estimations are typically revised upwards over time.

4 Revisions of the Structural Budget Balance: Time Series Approach

4.1 Data and Revision Size

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

The structural budget balances are calculated as described in the previous Section. The revisions in absolute terms are displayed in Figure 3. The average size of the first revision for Germany, Austria and France are 0.19 percentage points of potential output, 0.14 and 0.21, respectively, and thus somewhat larger compared to the production function approach. For Italy and the UK the first revision still amounts to 0.17 and 0.19 percentage points of potential output. Only for Spain, the revision size is with 0.18 percentage points smaller than in the production function approach. While the revision size diminishes over time, it still remains sizeable even after four years ($j = 8$). In contrast to the production function approach, however, the revision sizes of the time series approach decline faster and lie between 0.02 and 0.06 percentage points of potential output after eight revisions.

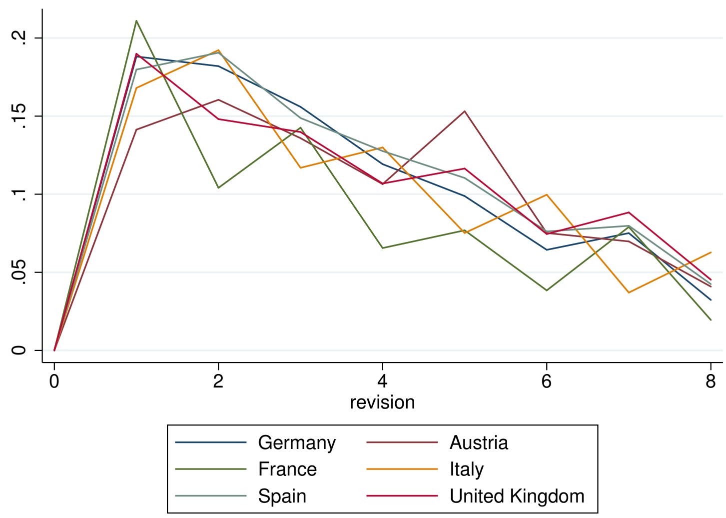

Figure 3: Size of the ex post revisions of the structural budget balance between two subsequent vintages using an HP filter (in percentage points)

4.2 Biasedness

The bias of the ex post revisions of the structural budget balance with respect to the initial estimate is shown in Figure 4. It was estimated using the same setting as in Section 3.2. For Italy, France and the United Kingdom, the bias is now smaller compared to the production function approach. In contrast to the production function approach, the downward revisions for Germany are not significant at a 5% level. Similar to the production function approach, there is no significant bias for Spain at a 5% level since the confidence bands are very large.

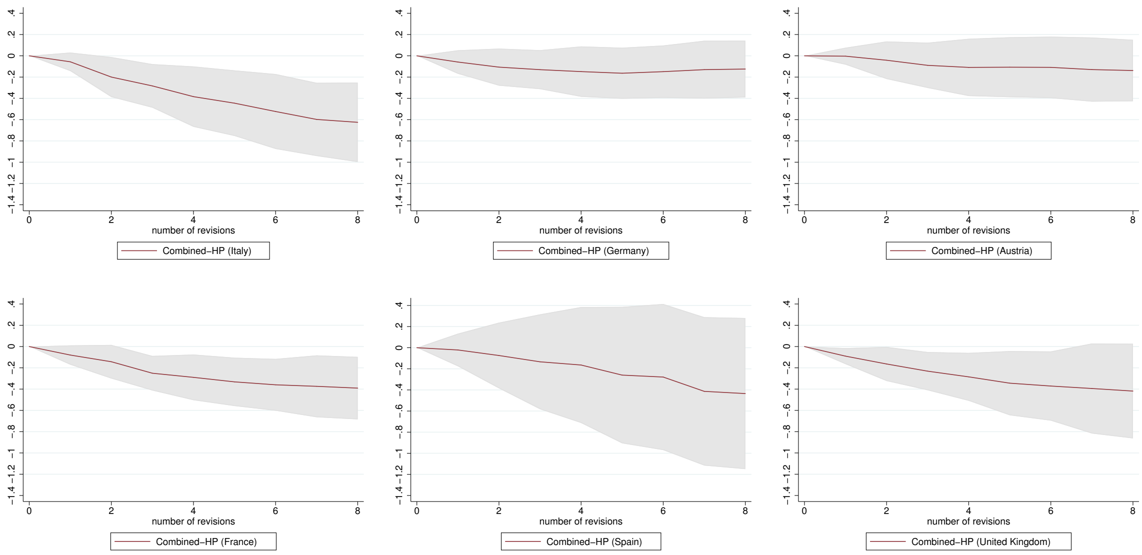

Figure 4: Bias of the ex post revisions of the structural budget balance with respect to the initial estimate using the an HP filter (in % of potential output)

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

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

5.1 Construction of the Survey-Based Structural Budget Balance

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

In many countries measures of capacity utilization are collected through business surveys. In the EU a representative sample of firms in services and manufacturing, which accounts for roughly 90% of gross value added, are asked for their capacity utilization on a quarterly basis.[7] According to the ECB (2007, p. 47, fn. 2), "survey respondents provide an answer about their overall resource utilisation, i.e. they consider both capital and labour inputs. This assumption is based on the overall content of the survey, which explicitly asks about various production constraints, including shortages of capital, labour and other inputs, suggesting that respondents have all those production inputs in mind when evaluating their capacity utilisation". The series for capacity utilization in manufacturing are available at least since 1996 and those for services at least since 2011 (see Table 1). Capacity utilization of the aggregate economy is the weighted average of both sectors.[8]

Since the degree of capacity utilization itself is not sufficient to identify an over- or under-utilization of the economy, a normal level has to be determined. Here we follow Marcellino and Musso (2011) and assume that an economy's normal level of capacity utilization is given by the mean of the time series.[9] Thus, for every country i and year t the capacity utilization gap

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

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

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

Table 1: First publication of capacity utilization data

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

Source: Eurostat.

($CUG_{i,t}$) is calculated as the deviation of the actual degree of capacity utilization ($CU_{i,t}$) from its mean ($\overline{CU_i}$):

$$CUG_{i,t} = CU_{i,t} - \overline{CU_i}. \tag{5}$$

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

The values for $CU_{i,t}$ are not revised after their initial publication. Thus, the only source of revision of $CUG_{i,t}$ is a change in $\overline{CU_i}$, which happens with every new observation. Obviously, a longer series with more observations has a more stable mean. As the series for capacity utilization in the service sector are relatively short, we use other survey data from the service sector to backcast them as long as possible and by this to enlarge the series and to stabilize the mean (see Appendix B for details). Figure 5 compares the latest estimate of the capacity utilization gap with the latest vintages (of spring 2018) of the output gaps estimated by the EC, the OECD and the IMF. For all countries, the capacity utilization gap is close to or within the range of these estimates represented by the shaded area. Thus, the production function-based output gaps and the capacity utilization gap are highly correlated. The correlation coefficients between the latest vintage of the capacity utilization gap and the output gap resulting from the production function approaches of the EC, the OECD and the IMF for the complete panel of countries ranges between 0.55 and 0.71 (see Table 2). The correlations with the output gaps of the EC are higher than with those of the OECD or the IMF. The correlations are substantially lower for the UK than for the other countries.

Table 2: Correlation between latest vintage for capacity utilization and latest vintages of latest vintage of traditional output gaps

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

capacity utilization as proposed in Marcellino and Musso (2011) or suggested in Orphanides *et al.* (2000) further decreases the revision size substantially as we will demonstrate.

Figure 5: Capacity utilization gaps (red line) and range of latest vintages of output gaps (in % of potential output or percentage point deviation from the mean)

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

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

$$\epsilon_{B,CUG} = \epsilon_{B,GAP}\epsilon_{GAP,CUG}. \tag{6}$$

While for $\epsilon_{B,GAP}$ we use the values provided by the EC (see Section 3.1), the semi-elasticity of the output gap with respect to the capacity utilization gap is defined as

$$\epsilon_{GAP,CUG} = dln(GAP + 1)/dln(CUG + 1) \tag{7}$$

and estimated via

$$ln(GAP_{i,t}) = \beta ln(CUG_{i,t}) + u_{i,t}. \tag{8}$$

For every country i the results for equation (8) are summarized in Table 3. As already indicated by the high correlation between GAP and CUG in Figure 5, the estimates for β are significant at the 1% level for all countries. Also the R^2 are high enough to justify the construction of the semi-elasticity this way, especially when comparing it to the goodness-of-fit of the estimations of the semi-elasticities in the production function approach (e.g. as in Price *et al.*, 2014). The estimates for β are then plugged into equation (6) as values for $\epsilon_{GAP,CUG}$. Finally, the survey-

Table 3: Results for equation (8)

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

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

based structural budget balance is computed by adjusting equations (1) and (2) to

$$X = B - \epsilon_{B,CUG} CUG - T. \tag{9}$$

Also the survey-based structural budget balance is expressed in % of potential output. By using this most common scaling we simplify the comparability of the methods and the implementation of the new approach.[10]

Figure 6 compares the survey-based structural budget balance with those resulting from the production function approach using the vintages of spring 2018 for each method. For all approaches we subtracted the one-offs published by the EC, which are available since 2010. This enables us to focus our analysis on the differences that result from the estimation of the cyclical components. Prior to 2010 we only display the cyclically-adjusted balance, which is identical to the structural budget balance if one-offs are ignored. For most of the period the results of both approaches are quite similar. However, at the end of the sample the survey-based structural budget balance is smaller for almost all countries. Thus, for a given budget balance the fiscal space turns out to be tighter in the survey-based approach than in the production function approaches. However, as we already know from Section 3 the initial estimates of the structural budget balances resulting from the production function approaches are systematically revised downward in subsequent estimates. Thus, in a few years it is very likely that the discrepancy at today's end of the sample will disappear, provided the initial estimates of the survey-based approach are at least less biased.

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

Figure 6: Structural and cyclically-adjusted budget balance: Survey-based (continuous and dashed red line) versus range of latest vintages from production function approach (in bn. euros)

5.2 Evaluation of the Survey-Based Approach

For estimating the bias of the survey-based structural budget balance, we re-estimate equation (4). The results are shown in Figure 7. The red line plots the estimates of α against the number of revisions $j-1$. The shaded area around the red line is the 95% confidence interval. For Germany, Austria and France the estimated structural budget balances are unbiased at the 5% level over all revision horizons. For Italy and the United Kingdom the estimates are unbiased up to the third revision. Then, the structural budget balance is systematically revised downward, which after 8 revisions amounts to 0.1 percentage points compared to the initial estimate. The results for Spain show a similar picture. Only the size of the downward revision is with 0.2 percentage points larger. A comparison with the estimated biases in both the production function approach (Section 3) and the time series approach (Section 4) however reveals that those biases in the survey-based approach which are statistically significant are quantitatively much smaller.[11]

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

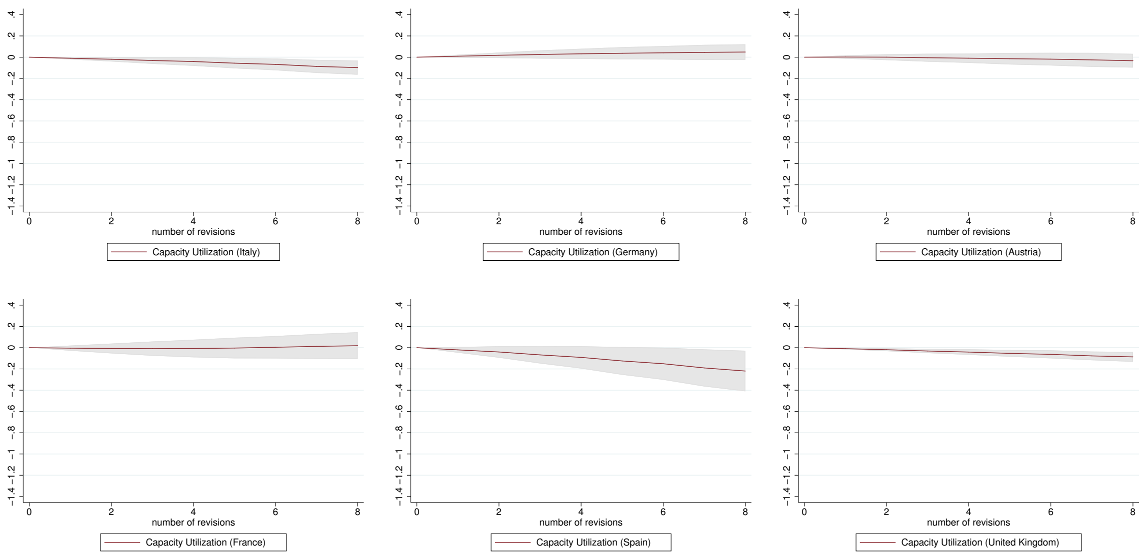

Figure 7: Bias of the ex post revisions of the structural budget with respect to the initial estimate using the survey-based approach (in % of potential output)

The tight confidence intervals in Figure 7 already hint at a small revision size of the survey-based structural budget balance. This result is supported by the direct calculation of the size of the ex post revisions as shown in Figure 8. The revision size is 0.02 percentage points of potential output for Austria, Germany and Italy and 0.1 percentage points of potential output for the UK at the first revision. Even in the case of Spain and France, where the revision size turns out to be largest, the survey-based approach performs better than both the production function approach and the time series approach. It declined from 0.28 percentage points (production function approach) and 0.18 percentage points (time series approach) to 0.03 percentage points for the first revision for Spain. For France, it declined from 0.20 percentage points (production function approach) and 0.21 percentage points (time series approach) to 0.03 percentage points for the first revision. This small revision size in the case of Spain in the survey-based approach comes at the cost of having a statistically significant bias at a 5% level. However, we believe that this cost is acceptable since the large revision size in the mainstream approaches for Spain made them hardly usable at all. In addition, this bias will vanish when the capacity utilization series will be long enough. In the case of the United Kingdom and Italy this will happen even sooner. Furthermore, it has to be emphasized, that this bias is merely a statistical one. Quantitatively, it is negligible.[12]

[12] An alternative interpretation of Figure 7 is the decline of capacity utilization during the Great Recession. That lead to a successive decline in $\overline{CU_i}$ in real time due to the small sample length.

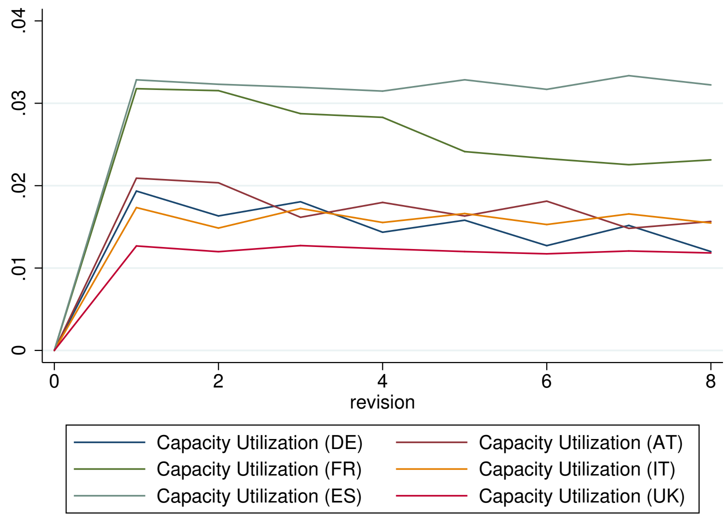

Figure 8: Size of the ex post revisions of the structural budget balance between two subsequent vintages using the survey-based approach (in percentage points of potential output)

6 Conclusions

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

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

The survey-based structural budget balances can be used to improve fiscal surveillance in the EU in general, but in particular the evaluation of compliance with EU fiscal rules. Currently, the assessment of compliance with those goals is based on an upward-biased estimation as shown in this paper. The survey-based approach would provide an unbiased (or at least much less biased) measuring tool for the concept of structural budget balances and their ex post evaluation. But also for the budgetary goals, which are formulated ex ante our method can be of use. At the end of the first month of a quarter, the capacity utilization of the respective quarter is available. The missing quarters can be estimated via standard forecasting methods already in use in various institutions. This would further help to improve the current evaluation of the fiscal rules since the estimation of output gaps (and consequently the estimation of structural budget balances) is also biased when assessing ex ante formulated output gaps (Kempkes, 2014).[13] We do not want to provide a forecasting method at this point. This is left for future research. Moreover, some parts of the preventive arm depend on the current position in the business cycle that is assessed by the output gap. Here, the mainstream output gap could be replaced by the survey-based capacity utilization gap.

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

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

References

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

BUTI, M., CARNOT, N., HRISTOV, A., MCMORROW, K., ROEGER, W. and VANDERMEULEN, V. (2019). Potential output and EU fiscal surveillance . https://voxeu.org/article/potential-output-and-eu-fiscal-surveillance, [Online, accessed 25-September-2019].

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

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

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

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

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

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

FEDELINO, A., IVANOVA, A. and HORTON, M. (2009). *Computing Cyclically Adjusted Balances and Automatic Stabilizers*. Technical notes and manuals, IMF.

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

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

HAVIK, K., MCMORROW, K., ORLANDI, F., PLANAS, C., RACIBORSKI, R., ROEGER, W., ROSSI, A., THUM-THYSEN, A. and VANDERMEULEN, V. (2014). *The Production Function Methodology for Calculating Potential Growth Rates and Output Gaps*. European Economy Economic Paper 535-November 2014, European Commission.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Silva, D., Monteiro, C., Leal, J. and Gutiérrez, E. (2016). *A Survey-Based measure of Output Gap for Portugal*. Artigos da autoria do GPEARI 06/2016, Gabinete de Planeamento, Estratéia, Avaliação e Relações Internacionais, Ministério das Finanças.

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

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

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

A Additional Analysis

Revisions of the structural budget balance can be mainly attributed to two sources: the budget balance B and its correction for the business cycle, the cyclical component $\epsilon_{BC,OG}OG$. In Section A.1, we show the revision size of the budget balance and compare it with the revision size of the cyclical component as shown in Figure 1. In Section A.2, we compare the relative revisions of the cyclical component with key figures of the System of National Accounts (SNA) to facilitate a classification of the revision size of the cyclical component.

A.1 Absolute Revisions

The revisions of the budget balance B and the cyclical component $\epsilon_{BC,OG}OG$ are directly comparable although the former is in relation to GDP and the latter in relation to potential output. Equation (1) shows that a revision of the budget balance B of the amount z to $B+z$ influences the structural budget balance X to the same amount as a revision of $\epsilon_{BC,OG}OG$ of the amount z to $\epsilon_{BC,OG}OG+z$. Figure 9 shows the average size of the absolute value of an ex post revision of the budget balance between two subsequent vintages.[14] It contributes to the revisions of the structural budget balance to a somewhat smaller amount than the revisions of the cyclical component $\epsilon_{BC,OG}OG$ as shown in Figure 1 in Section 3.1. However, the effect of these revisions of B on the structural budget balance X are also non-negligible. A possible reason for such large revisions at later stages might be changes in the System of National Accounts (SNA).

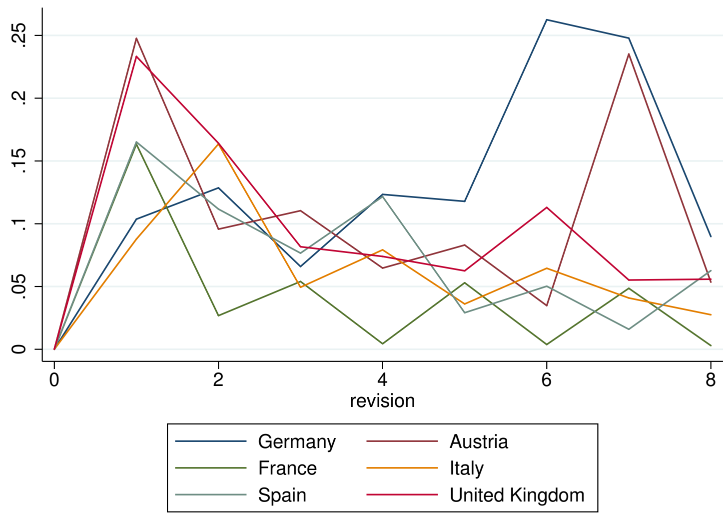

Figure 9: Size of the ex post revisions of the budget balance (relative to nominal GDP) between two subsequent vintages (in percentage points of GDP)

A.2 Relative Revisions

Key figures of the SNA have a different unit than the cyclical component. Thus, we need to compare their relative revisions instead of the absolute ones. The relative absolute revision of a

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

variable var is defined as $|(var_{t,i+1} - var_{t,i})/var_{t,i}|$. For the relative revision of GDP growth, we used the growth rates of GDP for the respective year: GDP_t/GDP_{t-1}. Figures 10 to 15 show that the relative revisions of the structural budget balances (production function and HP filter) are larger than the relative revisions of GDP (OECD and IMF), nominal GDP and the budget balance (relative to nominal GDP).[15] The HP filter series has a severe outlier.

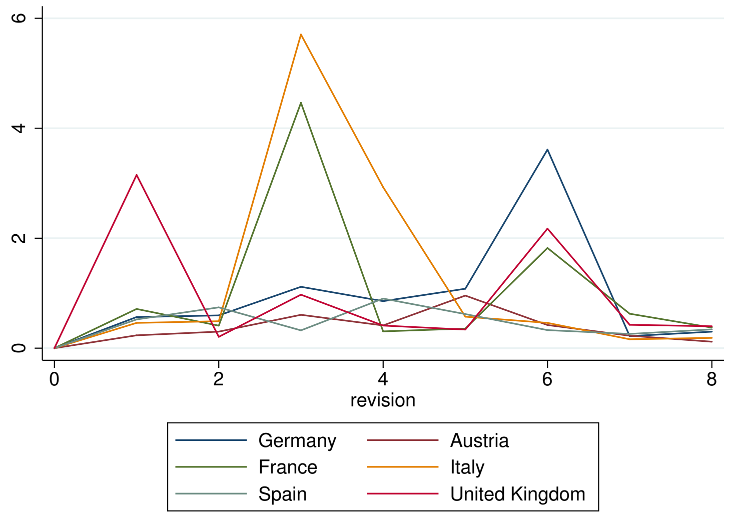

Figure 10: Relative size of the ex post revisions of the structural budget balance between two subsequent vintages using the production function approach (in percentage to previous vintage)

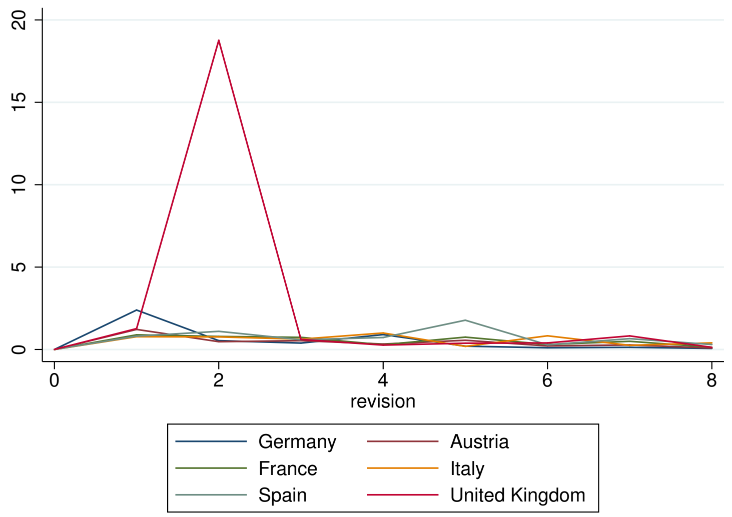

Figure 11: Relative size of the ex post revisions of the structural budget balance between two subsequent vintages using an HP filter (in percentage to previous vintage)

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

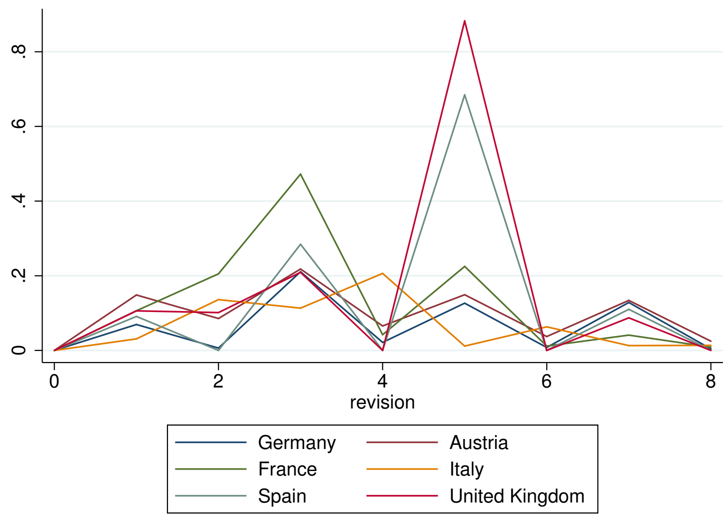

Figure 12: Relative size of the ex post revisions of the growth of GDP (OECD) between two subsequent vintages (in percentage to previous vintage)

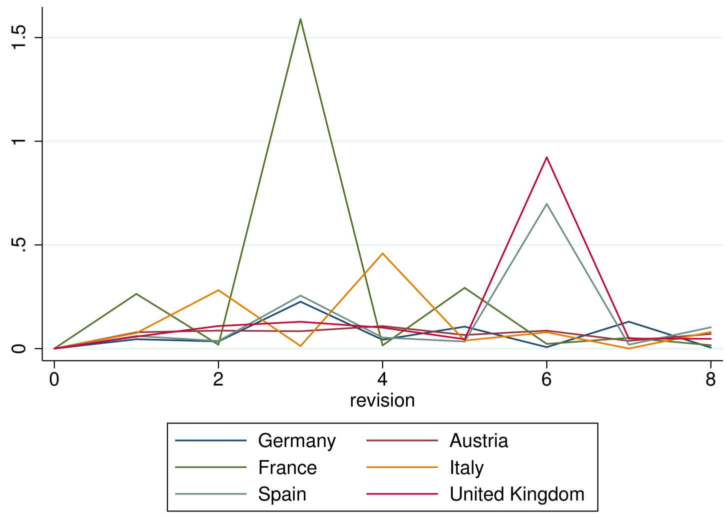

Figure 13: Relative size of the ex post revisions of the growth of GDP (IMF) between two subsequent vintages (in percentage to previous vintage)

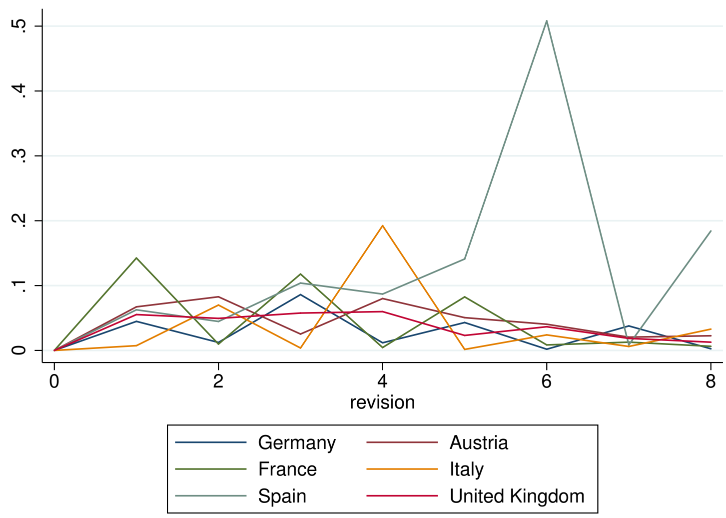

Figure 14: Relative size of the ex post revisions of the growth of nominal GDP between two subsequent vintages (in percentage to previous vintage)

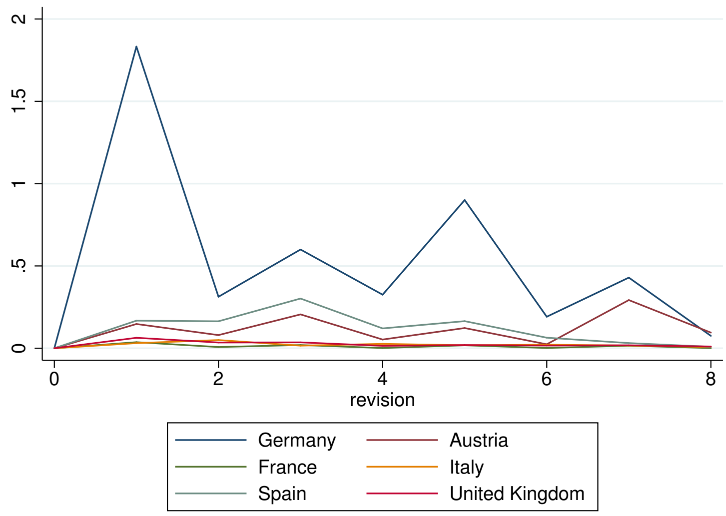

Figure 15: Relative size of the ex post revisions of the budget balance (relative to nominal GDP) between two subsequent vintages (in percentage to previous vintage)

B Data Adjustment

B.1 Backcasting Capacity Utilization in the Service Sector

For backcasting the series for capacity utilization in the service sector, we follow the approach suggested by Wohlrabe and Wollmershäuser (2017). They use a service sector confidence indicator that was already collected many years before capacity utilization in the European Commission's business surveys and which has a high correlation with the business cycle (see Table 4 for the availability of the series in the Eurostat database).

Table 4: First publication of confidence indicator in manufacturing and services

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

For backcasting capacity utilization we run the following regression

$$cu_{ser_{i,q}} = \alpha + \beta_i ci_{ser_{i,q}} + \epsilon_{i,q} \tag{10}$$

for every country i and quarter q, where cu_{ser} is the seasonally adjusted capacity utilization in the service sector, and ci_{ser} the seasonally adjusted services confidence indicator. In addition to country-specific analyses we also run a panel-regression.[16] The results for both, the panel and the country-specific regressions are summarized in Table 5. All estimates for β_i are significant at the 1%-level and the R^2 range between 0.13 and 0.67.[17] For the sake of simplicity we use the panel estimate of β for backcasting capacity utilization with the values of the service sector confidence indicator. Figure 16 shows the results of the backcast and compares the predicted series with the actual ones.

Table 5: Results for equation (10)

	(1) Panel	(2) AT	(3) FR	(4) DE	(5) IT	(6) ES	(7) UK
α	87.870***	89.178***	91.004***	88.726***	87.471***	84.051***	87.525***
	(0.079)	(0.287)	(0.139)	(0.285)	(0.219)	(0.232)	(0.217)
β_i	0.094***	0.097***	0.083***	0.055***	0.109***	0.113***	0.061***
	(0.013)	(0.014)	(0.015)	(0.020)	(0.015)	(0.011)	(0.013)
obs.	192	31	30	32	37	31	31
R^2	0.499	0.516	0.355	0.129	0.416	0.673	0.336

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

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

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

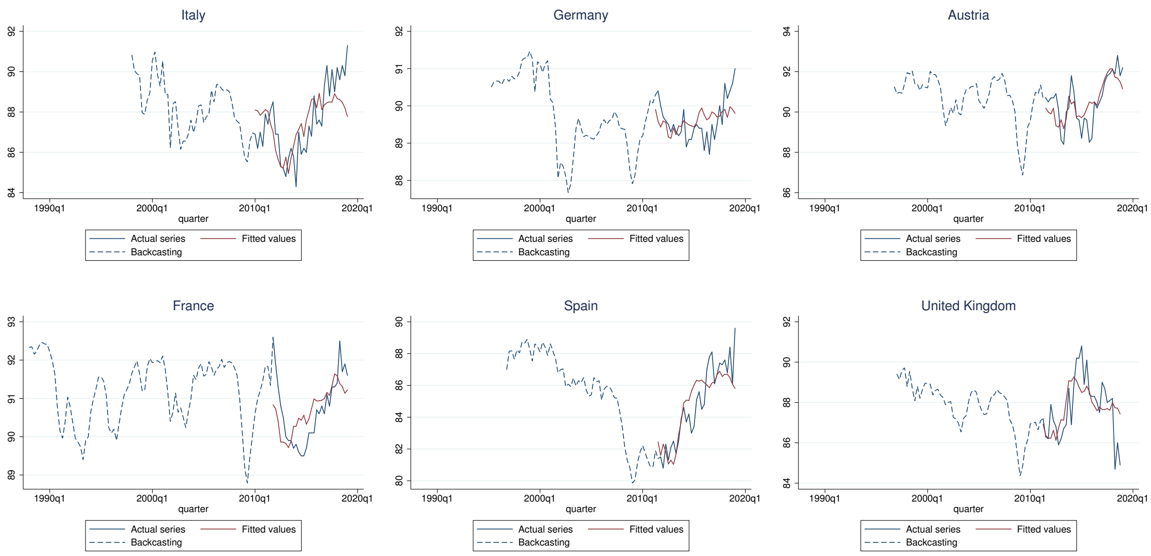

Figure 16: Comparison of actual capacity utilization series in services with the fitted values (in percentage points)

As a robustness check, we re-estimate equation (10) for the manufacturing sector:

$$cu_{man_{i,q}} = \alpha + \beta_i ci_{man_{i,q}} + \epsilon_{i,q}. \tag{11}$$

Results are summarized in Table 6. Figure 17 compares the predicted series with the actual ones.

Table 6: Results for equation (11)

	(1) Panel	(2) AT	(3) FR	(4) DE	(5) IT	(6) ES	(7) UK
α	82.053***	86.218***	84.444***	85.618***	76.236***	79.617***	81.785***
	(0.120)	(0.124)	(0.199)	(0.179)	(0.145)	(0.194)	(0.165)
β_i	0.200***	0.205***	0.158***	0.245***	0.249***	0.213***	0.150***
	(0.020)	(0.023)	(0.025)	(0.021)	(0.025)	(0.026)	(0.019)
obs.	745	93	113	137	137	128	137
R^2	0.496	0.634	0.364	0.641	0.540	0.422	0.427

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

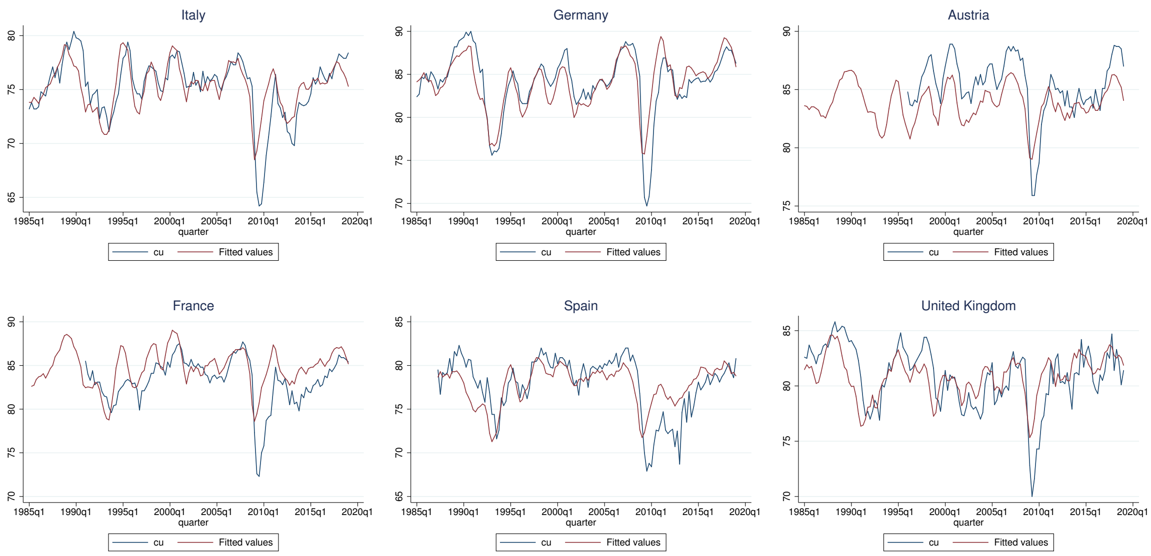

Figure 17: Comparison of actual capacity utilization series in manufacturing with the fitted values (in percentage points)

B.2 Enlarging Capacity Utilization of Whole Economy via Manufacturing

If the series for capacity utilization in manufacturing in a country dates further back than the backcasted series for capacity utilization in services we use the manufacturing series to enlarge the overall capacity utilization. Specifically we assume that the first difference of the capacity utilization in manufacturing and in the overall economy is the same. The high correlation between the capacity utilization in manufacturing and the overall economy, both in levels (see Table 7) and in first differences (see Table 8) supports this approach.[18] This enlarged series is only used to calculate the mean of the capacity utilization series for the overall economy. We enlarge the series for Italy back to 1988q1, for Germany back to 1991q1, for Austria back to 1996q1, for Spain back to 1988q1 and for the United Kingdom back to 1988q1.

Table 7: Correlation between capacity utilization in manufacturing and capacity utilization in whole economy

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

Table 8: Correlation between first-differences of capacity utilization in manufacturing and of capacity utilization in whole economy

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

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