

Moving in a Cycle: Are we going somewhere with Potential Output Estimations?

30th FMM Conference Contribution Paper

Leo Primavesi^{abcdef}

September 30, 2026

Abstract

Potential Output (PO) is a curious concept. It is central to the guidance of monetary and fiscal policy around the world, yet there is agreement across schools that we are far from providing a sensible measure for it. The obvious flaws of the PO methods that exist in political institutions are becoming harder to ignore even for mainstream proponents, a part of whom, especially since the permanent effects of the GFC, recognize the possibility of long-term effects on economic development and PO. Having started out on that assumption, Post-Keynesians reject a self-adjusting equilibrium at full sustainable utilization. In doing so, they disentangle two notions that jointly define PO in the Mainstream: the level of output, at which economic resources are fully utilized in a sustainable way under the given institutional setting is not necessarily coinciding with the highest inflation-stable level of output. This fact is often missed in work on PO, which leads to conceptual unclarity. The present work proposes that, to make progress in the ongoing debates around PO, one needs either to provide both a consistent view on what PO is and practically stable methods for its estimation, or reject the concept as an orientation tool for fiscal and monetary policy altogether and argue for alternative ways to guide these policies. This work tries to contribute to this process by evaluating two PO estimation methods discussed in recent literature - the Updated Okun Method and the Blanchard-Quah VAR model - on their ability to replace the methods currently employed in political institutions. The methods are both evaluated in terms of theoretical consistency and in terms of their revision behaviour by means of a real-time vintage data replication for Germany and Austria. The results show theoretical flaws as well as revisions that are large in magnitude, biased and partially pro-cyclical for both methods. The work closes with an outlook on how fiscal policy space estimation can function with and without utilizing the existing imperfect PO estimation methods.

^aEPOG Joint Master Program

^bSorbonne Université

^cConservatoire national des arts et métiers (cnam)

^dUniversité Sorbonne Paris Nord (USPN)

^eUniversité de Technologie de Compiègne (UTC)

^fHochschule für Wirtschaft und Recht Berlin (HWR Berlin)

Contents

List of Abbreviations	1
1 Introduction	3
2 Potential Output: Theory and Estimation	4
3 Econometric Methodology	8
3.1 Data	8
3.2 The Updated Okun Method (UOM)	9
3.3 The Blanchard-Quah SVAR Method (VAR)	14
4 Results	15
4.1 Okun Coefficients	15
4.2 PO Results	16
4.3 Revision Analysis	18
4.4 Evaluations	24
5 Discussion	25
5.1 Implications for Fiscal Policy	26
5.2 Limitations and Avenues for Future Research	27
6 Conclusion	28

List of Abbreviations

AIC Akaike Information Criterion

ARDL Autoregressive Distributed Lag

BIC Bayes Information Criterion

BP Bai-Perron

BQ Blanchard-Quah

CB Central Bank

CBO Congressional Budget Office

DSGE Dynamic-Stochastic General Equilibrium

EC European Commission

Fed US Federal Reserve Bank

GDP Gross Domestic Product

GFC Global Financial Crisis

IMF International Monetary Fund

NAIRU Non-Accelerating Inflation Rate of Unemployment

OECD Organisation for Economic Co-operation and Development

OG Output Gap

PK Post-Keynesian

PO Potential Output

RU-MIDAS Reverse Unrestricted Mixed Data Sampling

UOM Updated Okun Method

VAR Vector Autoregression

1 Introduction

The concepts of the Potential Output (PO) of an economy (and a corresponding Output Gap (OG)) are central in guiding monetary policy, fiscal policy and the design of structural reforms in the majority of central national and international economic institutions around the world. However, there is an array of problems associated with many common practices to theoretically conceptualize and empirically estimate PO. The theoretical assumptions, which the unobservable concept is based on, are far from uncontested (Horn et al. 2007, p. 11). Empirical estimation techniques are often unable to capture the complexity of the theoretical discourse and have an array of fundamental practical flaws ranging from substantial revisions of real-time estimations (Claeys et al. 2016) and pro-cyclical estimations (Heimberger and Kapeller 2017) to high sensitivity to arbitrary parameter choices (Fontanari et al. 2020) and “concerning epistemological perspectives” (Wang et al. 2009, p. 1). Adding to those flaws, an array of recent literature by economists of various backgrounds has found that estimation methods with modelling specifications and parameter choices *different* to the ones made in estimation methods commonly employed in (inter-)national institutions regularly reach significantly higher PO-Estimates (Palumbo 2015; Coibion, Gorodnichenko, et al. 2017; Jarociński and Lenza 2018; Fatás 2019; Heimberger 2020; Bassi 2024), fueling concern about a significant restrictive bias imposed on policy, e.g. by the common arbitrary modelling and parameter choices.

Coibion, Gorodnichenko, et al. (2017, p. 24) conclude that “[t]here is much work to be done to create better measures of potential GDP in real-time before policymakers rely on these too much”¹. Their call for further work on empirical PO estimation methods in order to provide a clearer guidance for economic policy that neither underestimates PO nor imposes a pro-cyclical bias on policy is echoed in other contemporaneous papers (Fatás 2019; Heimberger 2020).

This work aims to contribute to this research gap by evaluating two estimation methods that have recently been discussed as alternatives to the frequently used methods: The Updated Okun Method (UOM) proposed by Fontanari et al. (2020), who adapted the method by Okun (1962), and the Blanchard and Quah (1989) Vector Autoregression (VAR) model, reintroduced into the current discussion by Coibion, Gorodnichenko, et al. (2017). The central focus of the evaluation is an analysis *real-time estimate revision behaviour*, i.e. a comparison of a real-time ex ante and an ex post consideration with the same methodology at a later point in time, which is studied by means of an econometric vintage data procedure.

Section 2 briefly covers the state of the theoretical debate and introduces the most common estimation methods and their main problems. Section 3 then introduces the real-time data set as well as the replication processes for the UOM and VAR methods, which are applied for Germany and Austria. The econometric results of both the replication and the revision analysis are shown in Section 4. Besides discussions of the limitations and possible expansions of this work, Section 5 considers the implications of the sobering results on the ability of alternative methods for fiscal policy frameworks and asks the question whether the PO concept can meaningfully contribute at all to guiding policy in the future. Section 6 concludes.

¹ironically in reference to measures that policymakers rely on immensely across big economic institutions

2 Potential Output: Theory and Estimation

While there is no uniform definition, PO is most often defined as something along the lines of the “highest sustainable level of output” (Horn et al. 2007, p. 12), with ‘sustainable’ indicating that this level of production neither induces inflationary nor deflationary tendencies. Consequently, the concept most often includes some sort of Non-Accelerating Inflation Rate of Unemployment (NAIRU). Neither PO nor the NAIRU are empirically observable, they rather are attempts at a theoretical simplification of a more complex reality. Moreover, the theoretical properties of as well as assumptions about the concepts have not been uncontested historically and contemporaneously, which means that any attempt to evaluate the ability of PO estimation methods to provide meaningful results needs to consider a critical view on the theoretical notion of PO, asking the question of what it is precisely that we are trying to estimate². Horn et al. (2007, p. 11) describe that the debates around the theoretical concept of PO essentially run along the lines of three questions: (1) Does output (inherently) adjust to PO? (2) Is PO endogenous to actual output? and (3) Can gaps between PO and output persist over a longer time?

Based on the neoclassical idea that economies, under certain perfect conditions, tend towards an equilibrium with optimally utilized economic resources, the common answer to these questions in the macroeconomic mainstream³ and (inter)national institutions historically was (and, explicitly or implicitly, regularly still is), that output after a period of time *fully adjusts* to an *exogenous* PO. Hand in hand with this answer goes the empirical practice of equating PO with the permanent or trend component of actual observed output (with the OG being the cyclical component), a practice which is common to the point that renowned authors can introduce the two terms ‘potential’ and ‘trend’ output as synonyms (e.g. Orphanides and Norden 2002, p. 569).

The proposed channels for the full adjustment of output to PO however are strongly questioned: Real-Balance-Effects, which are assumed in some older New-Keynesian models as in Carlin and Soskice (1990), can be deemed implausible for various reasons (see e.g. Hein 2004, p. 46 or Horn et al. 2007, pp. 19–22 for a more detailed discussion). When proposing a Central Bank (CB) that follows a Taylor Rule as the mechanism for full adjustment, as suggested in New Consensus Models such as the ones discussed in Arestis and Sawyer (2002) or introduced in Carlin and Soskice (2009), one needs to assume both symmetric reactions and effects of interest rate changes (Hein 2004, p. 46). Especially whether this adjustment is sufficient to automatically fill the demand gap can be questioned when considering the zero lower bound as well as the possibility of a low or asymmetric interest elasticity of investment.

The question marks around the belief that economies automatically adjust after a demand shock were further enhanced by the empirical observation of a remarkable permanent slowdown effect of the Global Financial Crisis (GFC). Most established macro models, trusting in the aforementioned adjustment channels, could only explain the concurrent slowdowns in output and PO estimates with a

²For more comprehensive reviews of the theory debates around PO, see Horn et al. (2007) and Wang et al. (2009)

³I use the terminology of mainstream for what is described and closer examined as orthodoxy in Lavoie (2022, ch.1). This is also where a more precise demarcation of what I understand as Post-Keynesianism can be found.

supply-side shock unrelated to the GFC (Fontanari et al. 2020, p. 3), which seems highly improbable since this shock would need to be perfectly correlated with the GFC temporally (Krugman 2018) and in intensity across regions (Ball 2014).

The alternative explanation would be phenomena of *hysteresis*, i.e. the idea that a temporary demand shock can have permanent effects on the supply conditions and long-term growth, implying both the rejection of a full adjustment of output to PO and the endogeneity of PO to output. Already a debate in parts of the mainstream decades ago (e.g. Blanchard and Summers 1986; Topel 1990), the debate on hysteresis has gotten more attention after the GFC and still develops. While Coibion, Austin, et al. (2018, p. 7) describe the evidence for long-term effects of aggregate demand to be “tentative at most”, the consequences of the GFC also have resulted in calls for a profoundly changed macroeconomic framework and an adapted approach in macroeconomic policy as by Blanchard and Summers (2017), Fatás (2019) or Cerra, Fatás, et al. (2023), who call for fiscal policy to take a far more aggressive expansive stance and potentially take over the stabilizing role from monetary policy, which is seen as constricted by the zero lower bound.

The demand for a more activist fiscal policy to steer economic development has been issued for decades by proponents of the Post-Keynesian (PK) school of thought, which assume demand-led growth as a fundamental framework. Assuming aggregate demand to affect the long-term growth potential, PKs understand PO as *endogenous* and provide an array of possible channels for Hysteresis: In terms of endogeneity of the long run rate of growth, increased output can increase the working force as well as labour productivity growth through multiple channels, the latter of which is known as “Kaldor-Verdoorn-Effect” (León-Ledesma and Thirlwall 2002, p. 445). Further could a higher wage share, which goes along with higher output, push firms to increase labour productivity to lower labour intensity, which is known as the “Marx-Hicks” effect (e.g. introduced in Hein and Tarassow (2010, p. 11). Beyond this empirically well documented (Vogel 2009; Dray and Thirlwall 2011; Lanzafame 2010) endogeneity of the long run rate of growth, also the NAIRU, which codetermines PO, is seen as endogenous for various reasons listed in Hein (2023b, ch.5)⁴.

It is notable that when accepting the proposition that PO might be endogenous, it is valuable to reconsider the definition of PO as “the highest sustainable output level” (Horn et al. 2007, p. 12) with ‘sustainable’ defined as non-inflationary. If high demand affects PO positively, pushing output beyond it by demand stimulation might only lead to a short-term inflation and be sustainable in the long term. One consequence is to focus less on the inflation limit and rather on the limit of economic capacity when defining a sensible goal for demand stimulation. One example of this approach is Bassi (2024, p. 128), who simply defines PO through stating “Potential Output [...] implies the full utilization of the existing productive capacity”⁵, arguing that in the case of a low capacity utilization rate, under certain

⁴It is important to note that the very existence of a NAIRU is not uncontested in Post-Keynesianism. With different modelling assumptions, inflation either accelerates when passing an endogenous NAIRU which acts as a short-term inflation barrier (Rowthorn 1977; Hein 2023a), or simply rises with lower unemployment (Dutt 1987; Lavoie 2024). In the latter case, the concept of any point beyond which inflation accelerates or decelerates is rejected. The mention of a NAIRU in this work refers to the former view as presented in Hein (2023a)

⁵With maximum sustainable level of production as the so called *engineer-rated capacity* (Eichner 1976, p. 62) or *practical capacity* (Steindl 1952, p. 7)

conditions, extra stimulation could absorb excess capacity without creating persistent inflationary pressure (Bassi 2024, p. 129).

However, in a PK view, inflation can arise as a consequence of multiple forms of conflict (Hein 2024, p. 3) and does not necessarily imply fully utilized economic capacities. I propose that consequently, measures that estimate PO should clearly separate whether they are trying to estimate what I call “full-capacity PO”, the output level at which capacities are fully utilized (such as Bassi 2024) or “inflation-stable PO”, the level at which inflation is created. Many Measures which emerge out of a mainstream tradition at which PO represents an equilibrium state, in which those output levels are the same, include capacity indicators and inflation indicators in their methodology to estimate PO. Those measures might try to zoom in on two different output levels at the same time, assuming them to be the same, and end up not estimating anything meaningful at all.

This conceptual unclarity, however, is only one of many problems associated with the empirical estimation methodologies. I will give only a brief overview on the most important methods here, there are however more comprehensive literature reviews such as Cerra and Saxena (2000), Wang et al. (2009), and Horn et al. (2007) or Coibion, Gorodnichenko, et al. (2017). Mishkin (2007) defines three broad classes of PO estimation measures: First there are statistical methods, which impose little theoretical structure and just assume PO to be the low frequency variation in output time series. While they can range from univariate filters to more complex VAR models (such as the Fed model by Fleischman and Roberts (2011) or unobserved components models (e.g. Jarociński and Lenza 2018)), a common problem is that they restrict PO to be the trend component of output, often ruling out the possibility of endogeneity by default. The second group are production function methods, which conduct an independent estimate of an aggregate production function and plug in potential values of the relevant variables. Production function methods are the most common methodology in international institutions. The model which the European Commission uses to estimate the conjuncture component and consequently, fiscal space of EU-countries, follows a production function approach (Havik et al. 2014). Similarly the Congressional Budget Office (CBO) as well as the OECD and the IMF estimate PO based on a Cobb-Douglas production function including a NAIRU and filtered trend of input variables (CBO 2014; Coibion, Gorodnichenko, et al. 2017). The first main problem with this type of methodology can occur when potential values such as the NAIRU are just estimated by using statistical filters on unemployment, as is common practice e.g. in the EU methodology (Heimberger, Kapeller, and Schütz 2017). Assuming the NAIRU to be the trend of unemployment of course has similar problems as assuming PO to be the trend of output. The second problem is the inherent dysfunctionality of an aggregate production function with neoclassical assumptions such as the Cobb-Douglas production function (for a summary see Felipe and McCombie 2014 or Primavesi 2024)⁶. The third category are DSGE-Model based approaches. As apparent from the name, they consist of calibrating a structural DSGE-Model for the relevant economy with replicating actual output in order to identify the shocks that presumably hit the economy. PO then is the result of turning certain frictions and shocks off. Coibion, Gorodnichenko, et al. (2017, p. 5) point out that those estimations

⁶this problem can of course be overcome by using a different type of production function, such as the Leontief production function in Bassi (2024)

heavily rely on the estimation of a sophisticated model and the associated assumptions.

An intermediate conclusion seems dispiriting. A majority of methods assume either PO to be the trend component of output, which is simply not an acceptable assumption when taking the state of theory seriously, or other potential variables to be the trend component of their real counterparts, and there are hardly any serious attempts to introduce endogeneity in the model structures. Further, the institutional practices of calculating and applying PO estimates are intransparent as for multiple international institutions, it is not completely public how they arrive at their estimates. For it's Greenbook estimates, the Fed takes the approach to calculate the average of an array of different methods including statistical filters as well as more structural model-based procedures, weighted by personal judgment of the responsible economists (Edge and Rudd 2016). For the IMF the methodology differs considerably between countries. Nearly 80% of desk economists in the IMF use personal judgment among other techniques and in some countries, particularly low-income countries with low data availability, PO numbers are just entered based on judgment (Resende 2014). Both the high level of personal judgment in the institutional implementation of PO estimation and the methods themselves, in which central parameters rely on the assessment of economists which value seems plausible to them, leave a lot of room for personal bias. While this is unavoidable given the lack of plausible mechanistic measures, the degree of personal judgment in PO estimates, which are regularly presented as objectively given, is rarely made clear and could explain the phenomenon that many recent alternative methodologies consistently get higher PO values (Coibion, Gorodnichenko, et al. 2017; Jarociński and Lenza 2018; Fontanari et al. 2020; Bassi 2024) with a negative bias of the economists making those judgment calls at international institutions.

Beyond this array of theoretical and institutional problems, there is a central practical problem associated with PO estimation, which is that nearly all PO estimation methods produce *significant revisions*, i.e. changes of PO estimates of a given point in time when estimated at a later point. High revisions strongly draw into question the forecast or real-time results and a method, which frequently have high revisions cannot be considered as a sensible policy tool, as stated by an array of authors (e.g. Proietti et al. 2020; Horn et al. 2007; Claeys et al. 2016). The reasons for revisions are either revisions in the data, which are prevalent for the data used to estimate PO (Orphanides and Norden 2002), or the effect of the extra information gained in the data added after the original estimation. While both are unavoidable as reasons for revisions, they can result in either very high or somewhat moderate revisions based on how well the method can deal with that problem. A good method should be comparably stable with respect to the revision of data and the expansion of the sample (Proietti et al. 2020). But not only the magnitude of revisions is a problem. A problem frequently pointed out is *pro-cyclicality* of revisions, i.e. revisions that are positively correlated with the cycle. Especially downwards revisions in the years after crisis are a common observation for the methods employed in institutions (Coibion, Gorodnichenko, et al. 2017). Multiple authors point out the danger of self-reinforcing loops in the case of pro-cyclicality due to too pessimistic policy reactions in crises and too optimistic policy reactions in booms (Heimberger 2020; Fatás 2019).

I consequently now move on to a comprehensive analysis of revision behaviour of two of the discussed methods, the UOM and the Blanchard-Quah VAR model. Those methodologies have been

chosen as interesting research objects since they (1) have been discussed in recent contributions to debates around alternative PO measures in the literature and (2) explicitly try to address the danger of structurally underestimating PO. For both methods, there are some question marks on whether they are presenting a consistent theoretical picture, which will be discussed in more detail in the theoretical evaluation in Section 4. Time and space constraints of this work limits this analysis to two methodologies, which leaves out the evaluation of a few notable possible contributions to the debate, e.g. Jarociński and Lenza (2018) and particularly Bassi (2024), the detailed evaluation of which I pass on to further research.

3 Econometric Methodology

3.1 Data

This part describes the econometric methodology for a revision analysis of the UOM and BQ-VAR method, which begins by replicating the two methods for Austria and Germany. We have already established that a part of PO revisions can be caused by the data itself being revised quite substantially. In order to paint a realistic picture of revision behaviour, a rigorous revision analysis needs to capture the data revisions. An array of authors therefore points to the necessity of using vintage data (Giannone et al. 2012; Federal Reserve Bank of Dallas et al. 2011; Orphanides and Norden 2002). Such data contains a time series (or vintage) for every month that replicates the time series of the variable that was available in that month. I use data from the OECD Real-Time database (OECD 2026), which is introduced in Federal Reserve Bank of Dallas et al. (2011). The data contains monthly vintages of GDP in quarterly frequency and the unemployment rate in monthly frequency for Germany and Austria. After truncating the data to omit the periods which have data gaps in some vintages as well as the vintages with too few periods, the Austrian data ranges from 1996Q1 to 2026Q2 with 325 vintages between April 1999 and April 2026 and the German data covers the time frame from 1991Q1-2026Q2 over 327 vintages between February 1999 and April 2026. The Austrian data was seasonally adjusted for a few of the early vintages, for which the data source reported that the data were unadjusted and for the German data, a few early missing observations for early quarters in some vintages were filled from close vintages in to extend the sample. Those adjustments are documented in more detail in the replication code, which can be found at Primavesi (2026).

I test unemployment in monthly frequency and quarterly averages as well as GDP for stationarity with the four standard unit root tests (ADF, PP, DF-GLS and KPSS). Table 1 shows the results for all variables over all vintages. For Germany, all variables are deemed stationary by at least three tests in all vintages at the 10 percent level. While for Austria, there are a few vintage editions for which only 2 or occasionally only 1 test shows stationarity at the 10 percent level, those are either very early editions or editions in the crisis years 2008 or 2020. Assuming that the problem why the tests do not show stationarity for those editions lies either in noisiness due to too few observations or in large outliers at the end of the sample, I do not take this as strong evidence for questioning the stationarity of the data-generating process. I therefore move on with the data as is and stay cautious for uncommon behaviour of the method in the non-stationary editions.

Table 1: Unit Root Test Results Across Vintage Editions

Tests passed	Austria			Germany		
	Δu_t^M	ΔGDP_t^Q	Δu_t^Q	Δu_t^M	ΔGDP_t^Q	Δu_t^Q
0/4 (non-stationary)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
1/4	1 (0.3%)	7 (2.2%)	7 (2.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
2/4	3 (0.9%)	43 (13.2%)	48 (14.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
3/4	23 (7.1%)	51 (15.7%)	37 (11.4%)	248 (75.8%)	54 (16.5%)	129 (39.4%)
4/4 (stationary)	300 (91.7%)	224 (68.9%)	233 (71.7%)	79 (24.2%)	273 (83.5%)	198 (60.6%)
<i>N</i> (editions)	327	325	325	327	327	327

Notes: Each cell reports the number of real-time vintage editions (and share) for which the given number of the four unit root tests (KPSS, ADF, PP, DF-GLS) indicate stationarity at the 10 percent level. Δu_t^M : monthly change in unemployment; ΔY_t^Q : demeaned quarterly GDP growth; Δu_t^Q : quarterly change in unemployment.

3.2 The Updated Okun Method (UOM)

The Updated Okun Method (UOM) does not really fit into the earlier introduced categories by Mishkin (2007). Inspired by the initial methodological proposal of Okun (1962), the method was introduced and expanded by Fontanari et al. (2020), Fontanari et al. (2021b), and Fontanari et al. (2021a). This method consists of connecting a target rate of unemployment with an estimate of the okun relationship between unemployment changes and output changes. Full employment output is then interpreted as the output that would be reached connected with the target rate of unemployment when a change in unemployment would change output according to the estimated okun coefficient. The normatively chosen target rate of unemployment is supposed to represent an inflation-stable high-activity output. Practically a historically low unemployment associated with low inflation is chosen as target unemployment.

The Replication of the UOM generally follows the methodology of Fontanari et al. (2020) as well as the implication procedure of Fellingner (2022) for Austria. I aim to replicate as precise as possible the steps that a researcher or statistician would take with to create a PO Estimate series under the constraint of accessing the data at a given vintage. In doing so, allowing for some reaction of the method to the changing data (i.e. adding a structural break variable in vintage editions that are late enough to observe a pattern in the data that hints towards a structural break) is necessary to mimic that a researcher at this point in time would have adapted the method to the data. However, if the method changes too much from edition to edition, the *revision analysis* becomes less powerful, since revisions due to method changes can make the results of revisions more about how much the method changed rather than about the intended evaluation of how much of a problem data revisions and sample extensions pose to any given PO estimation method. Replicating the UOM over all vintages is therefore not a trivial technical task but involves qualitative decisions and judgment calls, which are described and contextualized in this section.

The first central step is to estimate the *okun coefficient*, the relationship between unemployment and output. As in Fontanari et al. (2020), the variables entering the regressions are unemployment rate change and GDP growth. I estimate okun coefficients with multiple regression models. First, I attempt

to work with the unemployment data in monthly frequency. Since I need to regress a low-frequency variable on a high-frequency variable, I employ a reverse unrestricted MIDAS (RU-MIDAS) regression as introduced in Foroni et al. (2018) with the following specification:

$$\Delta u_\tau = \alpha + \sum_{k=0}^3 \beta_k \cdot y_{t(\tau)-k}^Q + \sum_{j=1}^p \gamma_j \cdot \Delta u_{\tau-j} + \varepsilon_\tau \quad (1)$$

$$\beta = \sum_{k=0}^3 \beta_k \quad (2)$$

with Δu_τ as the monthly change in the unemployment rate, $y_{t(\tau)}^Q$ is the quarterly GDP growth rate observed in the quarter $t(\tau)$ corresponding to month τ , ε_τ as the error term and p as the value in a range between 8 and 10 that yields the lowest AIC. The range as well as the set GDP growth lags have been determined by applying a Ljung-Box test to see in how many of the vintages there was a high probability for autocorrelated errors (which would signify that the regression misses information in the data) and by checking the plausibility of the resulting okun coefficients against values in the literature. Following Fontanari et al. (2021b), the okun coefficient β is understood as the cumulative dynamic effect of GDP on unemployment, which is the sum of the contemporaneous and the lagged effects. It might however be a valuable discussion to consider whether the long-run multiplier β_{LR} as defined in Equation 3, which includes the autoregressive effects of increased unemployment that follow an the effects in GDP, could be included in the okun coefficients.

$$\beta_{LR} = \frac{\sum_{k=0}^3 \beta_k}{1 - \sum_{j=1}^p \gamma_j} \quad (3)$$

I then switch to a quarter-on-quarter-framework with the quarterly averages of the unemployment rate, both for robustness and comparability with Fontanari et al. (2021b), who do not use monthly data. Here, I begin with a simple *Autoregressive Distributed Lag (ARDL)* model with quarter-on-quarter-frequency where the number of autoregressive unemployment lags p and the number of GDP lags q is defined by the Akaike Information Criterion (AIC):

$$\Delta u_t = \alpha + \sum_{k=0}^q \beta_k \cdot \Delta y_{t-k} + \sum_{j=1}^p \gamma_j \cdot \Delta u_{t-j} + \varepsilon_t \quad (4)$$

$$\beta = \sum_{k=0}^q \beta_k \quad (5)$$

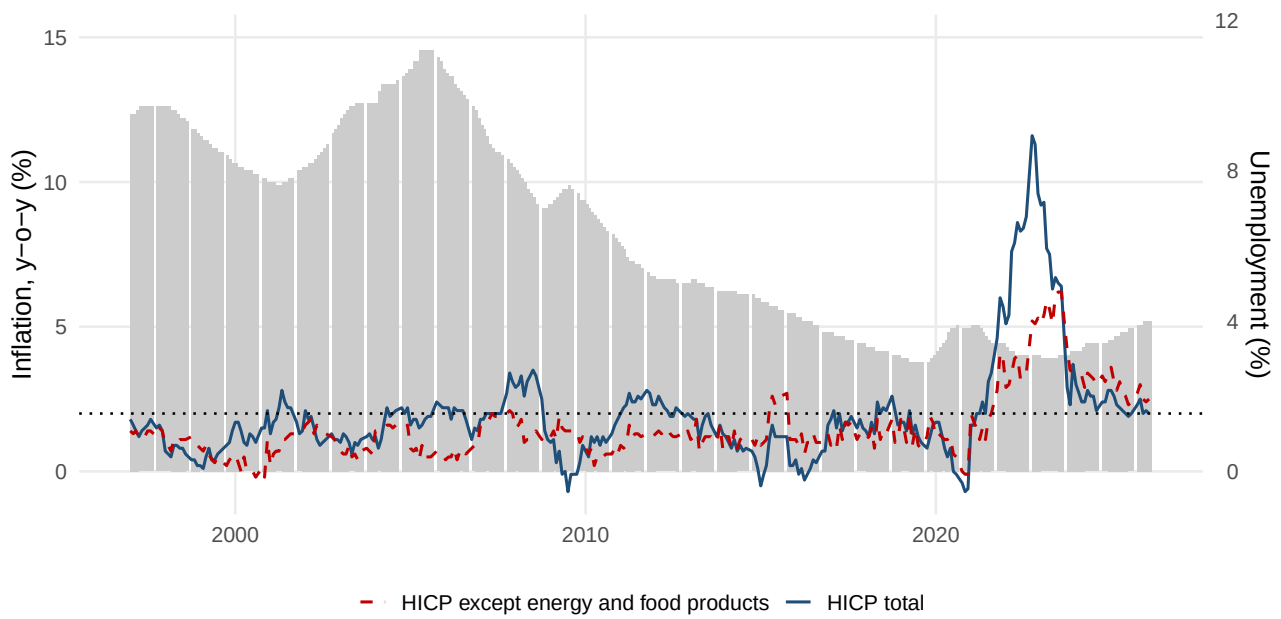
Again, the okun coefficient in every vintage is the sum of the contemporaneous and lagged coefficients. I then move to two procedures aimed at estimating structural breaks. First, I run a 50-quarter window rolling regression. The window size is chosen arbitrarily, but changing it does not significantly alter the range or general movement of coefficients. The sum of the GDP coefficients of each regression serves as the okun coefficient that is estimated for the last quarter of each regression. This yields a timeline of okun coefficients for each vintage, which is interpreted as the structural devel-

opment of the okun relationship when estimated in that vintage. The second procedure to identify breaks is a Bai-Perron (BP) test, as introduced by Bai and Perron (1998), which is designed both to find multiple breaks as well as, by means of the BIC, identify a sensible number of breaks to not miss structural changes while avoiding overfitting. To run the BP-test, I need to first assume a structure of lags. In order to not miss any breaks, I run the BP test first on the lag structure decided by the AIC separately in every vintage and then for the three options that were the most common lag structures over all vintages, assuming that those are models that might give sensible results for the underlying economies. Figure A1 shows the results for Germany and Austria. We can see that there are only few suggested breaks in Austria in general over the time frame. None of them show up consistently enough across vintages to be considered further, especially since all models that include autoregressive parts suggest nearly no breaks over all vintages and have rather low autoregression left in the residuals, which can be seen by the rates of vintages with rejected Ljung-Box tests. For Germany the picture is a bit more complicated. The high Ljung-Box rejection rates which stay prevalent in the ARDL models suggest the possibility of improving the model by including structural breaks. While a regression without lagged components suggests multiple structural breaks over nearly all vintages, when we test for breaks with an autoregressive model, only a few structural breaks remain. There are only two breaks which show up fairly consistently over vintages, a break around 2005Q3, which seems plausible given the implementation of the “Agenda 2010” as well as the close proximity to the GFC, as well as a break in 1993-1994, which shows up in the early vintages. Based on this information, I implement control variables for the breaks in a separate ARDL regression. I move on with including the 2005Q3 break in all vintages past its first appearance in ARDL models in the September 2012 vintage. While the break does not show up in the BP tests on the ARDL models after the February 2020 vintage, it does show up in a no-lag model and is economically plausible, which is why I assume that researchers could have grounds to include it even after 2020. The 1993-1994 break is also economically plausible given the *Treuhand* restructuring of the German Economy after the Reunification, however it is quite early in the sample, which starts in 1991Q1 and the okun coefficient for the first snippet of the sample might vary significantly when including it, which could affect the revision analysis. I consider a break model both with this break in the vintage time frame in which it shows up (Mar 2004 - May 2012) and without it. The other structural breaks that the no-lag regression model suggests are not considered further, since this is assumed to overfit the model and they do not show up when considering autoregressive components. Table 2 shows Ljung-Box rejection rates for different AIC search specifications with and without the 1994 break in Germany. The chosen specifications are a simple ARDL model with $p_{max} = q_{max} = 4$ for Austria and a break-augmented ARDL which includes the 2005 break as the only break with $p_{max} = q_{max} = 3$, which are the measures evaluated for the revision analysis. The rolling regression coefficients as well as RU-MIDAS coefficients are taken as robustness checks for the okun coefficients.

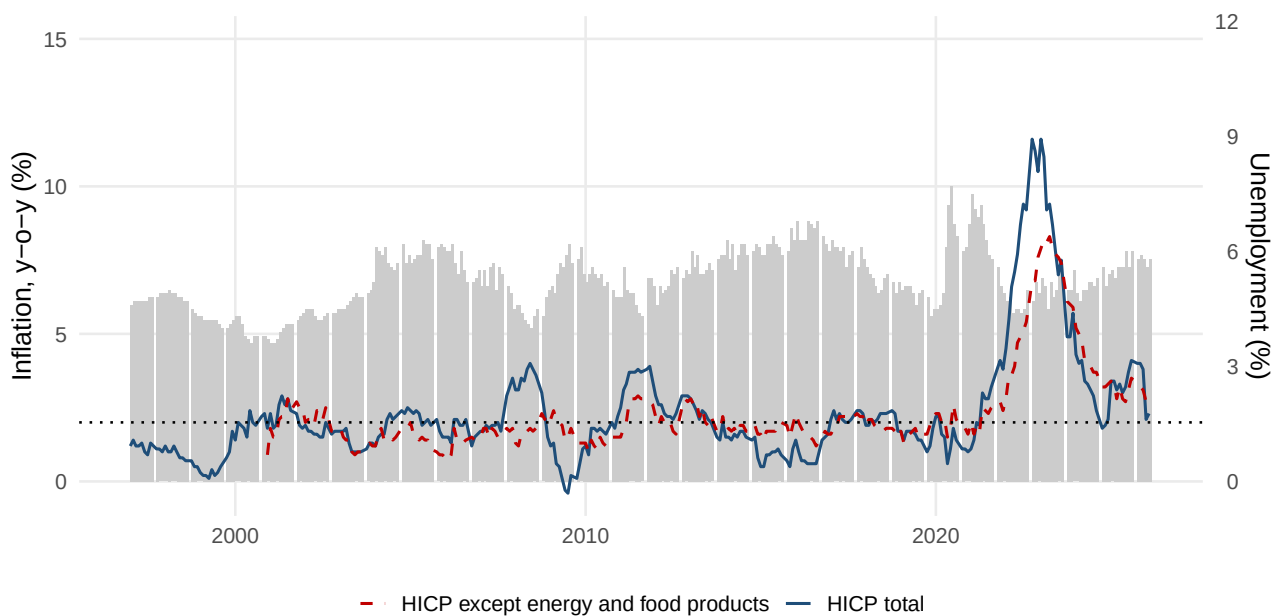
Table 2: Ljung-Box Rejection Rates — ARDL and Break-Augmented ARDL by Vintage; selected models in **boldface**

Country	Max lags	Model	1994 break	Rej. 5%	Rej. 1%
<i>Germany (DE), 326 vintages</i>					
	2	Pure ARDL	—	141 (43.3%)	88 (27.0%)
	2	Break model	included	140 (42.9%)	89 (27.3%)
	2	Break model	excluded	142 (43.6%)	90 (27.6%)
	3	Pure ARDL	—	114 (35.0%)	52 (16.0%)
	3	Break model	included	118 (36.2%)	52 (16.0%)
	3	Break model	excluded	110 (33.7%)	52 (16.0%)
	4	Pure ARDL	—	178 (54.6%)	118 (36.2%)
	4	Break model	included	174 (53.4%)	117 (35.9%)
	4	Break model	excluded	176 (54.0%)	122 (37.4%)
<i>Austria (AT), 324 vintages</i>					
	2	Pure ARDL	—	131 (40.4%)	100 (30.9%)
	3	Pure ARDL	—	137 (42.3%)	102 (31.5%)
	4	Pure ARDL	—	21 (6.5%)	2 (0.6%)

The second central question is the decision about the target unemployment rate. There are two quite simple approaches to this question in the literature. As discussed earlier, Okun (1962) just assumed 4% to be an inflation-stable high-demand unemployment. Fellingner (2022) follows this method by arbitrarily choosing a level that historically was associated with high demand and low inflation. Fontanari et al. (2020) in their work take the lowest unemployment rate observed up to that point in time. This, per definition, makes the resulting PO bigger than actual output, i.e. overutilization impossible, which is why I selected to apply fixed targets here.



(a) Inflation and unemployment for Germany



(b) Inflation and unemployment for Austria

Figure 1: Inflation and unemployment developments for Germany and Austria. Source: ECB (2026) (Inflation data), OECD (2026) (Unemployment, March 2026 vintage), own depiction

Figure 1 shows inflation and unemployment rate movements for Germany and Austria. For Germany (Figure 1a), in 2020, an unemployment rate of around 4% coincided with a low inflation, in Austria around 2002 4.5% of unemployment was possible with an inflation around the ECB target. To make sure the target is a high-activity target, I choose 3.5% for Germany and 4% for Austria.

Finally, PO and the OG are estimated according to the following formula:

$$Y_t^* = Y_t \left(1 - \frac{u_t - u^*}{100 \bar{\beta}_t} \right) \quad (6)$$

$$OG_t = \left(\frac{Y_t - Y_t^*}{Y_t^*} \right) \cdot 100 \quad (7)$$

with Y_t^* as UOM PO, OG_t as the output gap u_t as unemployment, u_t^* as the target unemployment rate, and $\bar{\beta}_t$ as the potentially time-varying okun coefficient. The okun coefficient is estimated anew for each vintage, which is to be taken account in the results of our revision analysis. A separate specification would be the trailing average over the last 12 vintage editions, following the idea of the okun relationship as a long-run structural relationship, for which researchers will always consider past estimations to a degree. This smoothing out of historical okun estimations would of course lower revisions over the first year, but would make some of the revision procedures impossible.

3.3 The Blanchard-Quah SVAR Method (VAR)

The second method that I analyse in this work is part of the *statistical methods* category: **the long-run restriction VAR model** by Blanchard and Quah (1989)(BQ), which has been reintroduced as a starting point for an alternative method by Coibion, Gorodnichenko, et al. (2017). The procedure consists of estimating a bivariate VAR of output growth and the unemployment rate. To replicate it, I estimate a bivariate VAR (p) which includes the demeaned unemployment rate and GDP growth variables. In the lag specification I diverge from BQ and Coibion, Gorodnichenko, et al. (2017), who simply estimate a VAR (8). I do so for two reasons. First, assuming a lag length is not consistent with the diagnostic of the information criteria. Second, especially for the Austrian data, a VAR(8) leads to a significant amount of vintages in which the model explodes. The lag specification I choose is based on the percentage of explosive models (i.e. models with at least one eigenvalue outside the unit circle) over all vintages. A1 shows the percentage of explosive models over all vintages with set different lag specifications. Looking to include the Information criteria for lag choice, reduce the percentage of explosive models and avoid including too few lags, we decide on $p_{max} = 4$ for Austria and $p_{max} = 8$ for Germany.

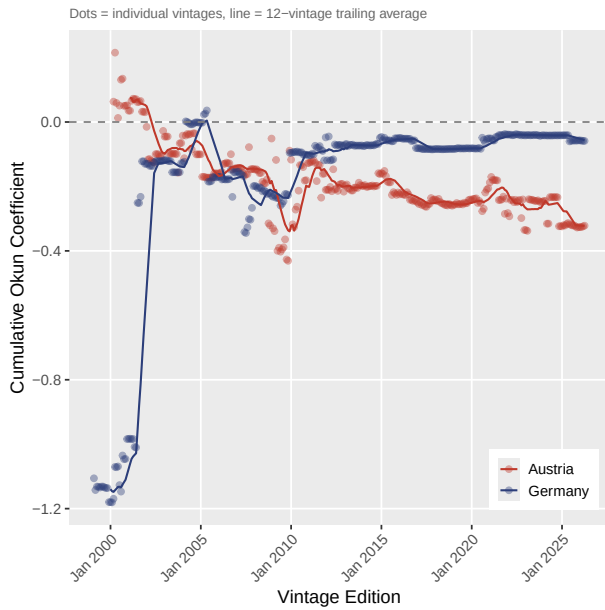
The structural VAR is then identified through a long-run restriction with the following assumptions: There are two types of shocks on output, permanent and temporary. Both are assumed to not have long-run effects on unemployment, but one is assumed to have long-run effects on output (permanent effect) while the other has not (temporary effect). The timeline of GDP only affected by permanent shocks then is interpreted as PO. New Keynesians, BQ label the permanent shocks *supply shocks* and the temporary shocks *demand shocks*. While this seems alarming from our theoretical perspective on the first glance, this is only a question of framing. Then, a counterfactual timeline is constructed in which only permanent shocks hit GDP and propagate. This timeline is interpreted as PO. It is important to note that, for the first p periods, since the VAR has no lags it cannot estimate a counterfactual timeline. Those periods therefore are pinned down by actual history and cannot be interpreted as PO. This then leaves us with a timeline of PO *changes*. If we want to compute OGs, however, we need

to get to a *level* measure of PO, which achieve by setting an anchor, one point in the sample at which it seems economically plausible that the output gap was zero. This would be a point in time with constant low inflation as well as low unemployment. Out of concern to make the revision analysis noisy, we want to pick the same anchor across all vintages, which means we need to drop vintages which are too early to contain the anchor point. Consequentially, the target anchor should be early in the sample. Considering Figure 1b again, we find a plausible anchor in Austria in 2001Q2. For Germany, 2019Q2 would provide a plausible anchor point. This is, however, very late in the sample and would drop too many vintages for the revision analysis. We therefore decide for 2001Q2 as well, which marks an early relative low point in unemployment while inflation is stable which, as we will see, has significant consequences in levels and direction of output gaps for the results.

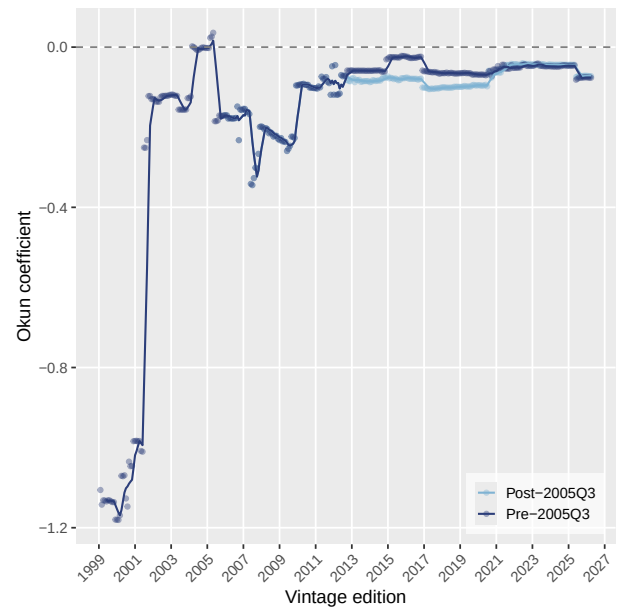
4 Results

4.1 Okun Coefficients

The results for the linear and break-augmented ARDL and of okun coefficients across vintages can be found in Figure 2. After a brief period of very low okun coefficients in the early vintages, which can be attributed to a low sample size, the okun coefficient stabilizes between 0 and -0.1 in Germany. For Austria, the early editions are implausibly positive in the early vintages and stabilize between -0.1 and -0.3, tending towards the latter in later vintages. These results are consistent with the estimations of the RU-MIDAS regression as well as the rolling regression, which are shown by Figures A2 and A3. The Austrian values as well as their development over time are roughly in line with Fontanari et al. (2021b) as well as Fellingner (2022). Interestingly, my coefficients show a smaller relationship than the ones of Fontanari et al. (2020). Looking into further literature, the results of Dunsch (2016) also okun coefficients similar to those of Fontanari et al. (2021b), while Nebot et al. (2019) find an asymmetric reaction with okun coefficients of zero in positive GDP growth rates and an okun coefficient at around -0.1 for negative growth rates, in line with the results presented here.



(a) Cumulative Okun Coefficients estimated by ARDL-Model, lag selection by AIC



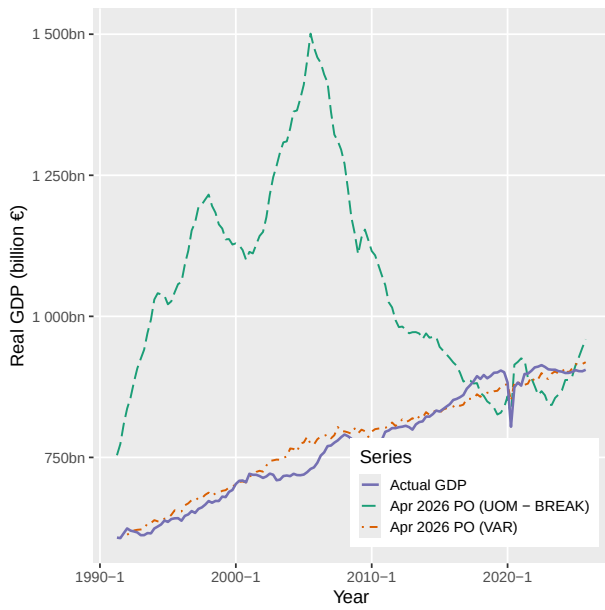
(b) Break-Augmented Okun Coefficients, DE

Figure 2: Okun coefficient with and without breaks

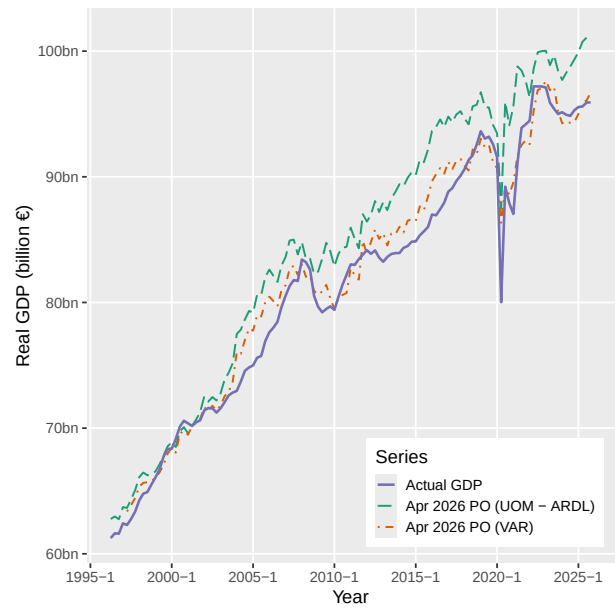
4.2 PO Results

Figure 3 shows the PO timeline for the most recent vintage as estimated by the Break-Augmented UOM as well as the VAR for Austria and Germany respectively.

Most notably are the UOM results for Germany, which are economically implausible over large part of the sample. In those phases, the OG turns widely positive, signalling a level of underutilization (or distance to an inflation-barrier depending on which PO concept is employed) that is simply unrealistic. This is the consequence of the earlier discussed low okun coefficients in combination with high unemployment (i.e. a large deviation from the set unemployment target) and therefore an expected method failure. Fontanari et al. (2020) themselves state that the UOM relies on the existence of a sufficiently strong negative okun relationship, which we simply do not observe in the data for Germany. This (non-)observation is consistent across estimation methods, vintages and can be found in parts of the literature such as Nebot et al. (2019).



(a) Germany

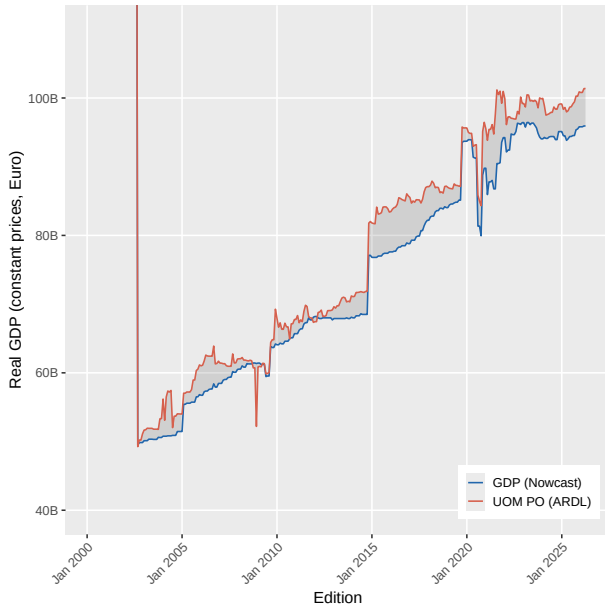


(b) Austria

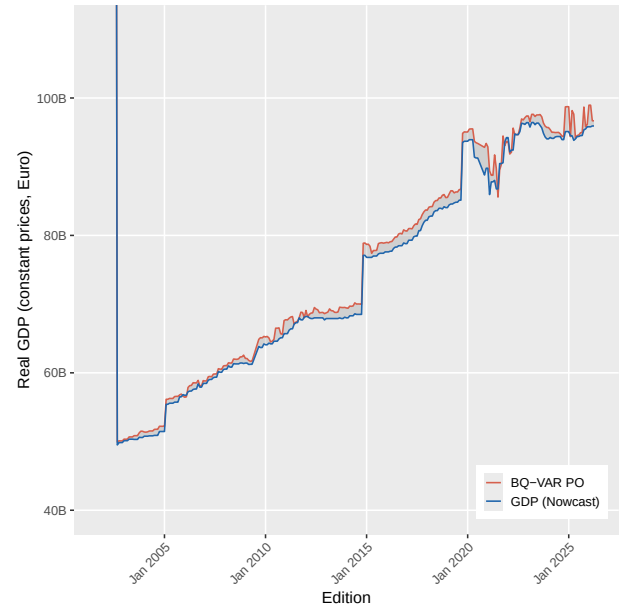
Figure 3: Potential Output Estimations by VAR and UOM for Germany and Austria, March 2026 Vintage

The VAR results shows Germany's output to fluctuate around a steady PO. However, this result significantly depends on the choice of anchor period. Choosing a different anchor period leads to a simple vertical movement of the entire PO line. If a quarter in 2019 had been chosen, PO would have been above output for the entire sample. This would seem to be the more plausible timeline, also when considering the inflation data, which does not seem to replicate the overutilization that the VAR model assumes. We will note that a timeline of underutilization over the entire time frame as a consequence of a different anchor period is the more plausible VAR result. We will, however, move on with the 2001Q2 anchor for the revision analysis. Since a different anchor only means a vertical movement, which is most probably of similar magnitude in all vintages, this should not affect the revision results significantly.

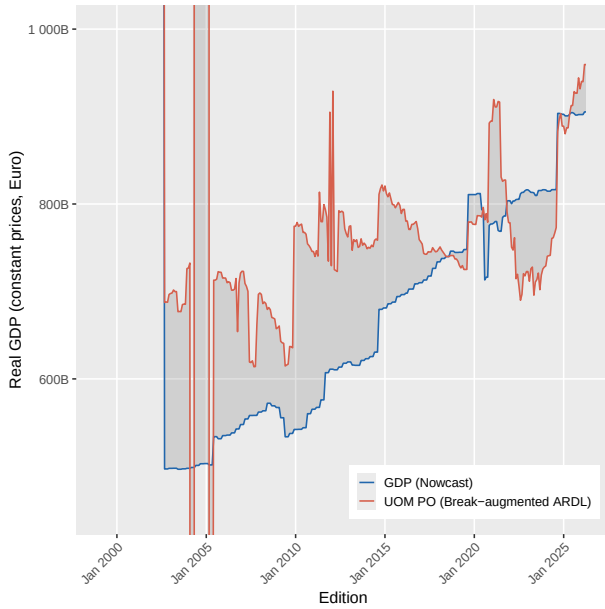
In Austria, due to a sufficiently strong okun relationship, we arrive at more plausible results for the UOM. We find Austria to be consistently below PO by means of both the UOM and the VAR measure.



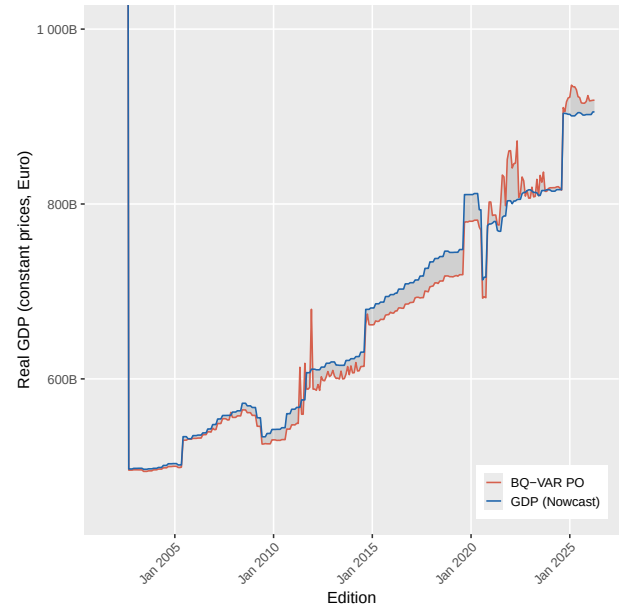
(a) UOM (ARDL), Austria



(b) VAR, Austria



(c) UOM (Break-augmented ARDL), Germany



(d) VAR, Germany

Figure 4: Latest PO estimates at each vintage

4.3 Revision Analysis

Figure 4 show the historical series of PO nowcasts (i.e. in this context the latest PO estimation at every vintage), which is the closest we come to the measure policy makers have to rely on⁷. This

⁷What we observe is the PO estimation at the first point where official data for GDP and unemployment are available which is around a quarter after the estimated point in time. Since the focus of the analysis is the ability of the measures to inform fiscal policy, which takes some time to initially react, this is deemed sufficient. An estimation of the nowcasts that researchers would be able to make would require a vintage set of flash estimates and an array of further analyses, which is far beyond the scope of this work

series, or rather the OGs it produces, are therefore the basis for the analysed revisions $R_t^{(x)}$:

$$R_t^{(x)} = \hat{OG}_t^{(0)} - \hat{OG}_t^{(x)} \quad (8)$$

with $\hat{OG}_t^{(0)}$ as the initially (at vintage 0) observed output gap for quarter t , and $\hat{OG}_t^{(x)}$ as the OG for the same quarter x vintage editions later. We conduct the revision analysis on OGs rather than PO because there are multiple estimation differences between vintage editions that change the *level* of PO and GDP likewise but should not affect the calculated output gap, such as deflator rebasing procedures over years, which can be seen in the regular jumps in all subplots of Figure 4. Further notable are some implausible jumps in the estimates as well as the data in the early vintages. The strong drop in output and PO between the Jan 2002 and Feb 2002 vintages, which can be observed for all methods, is a result of the statistical rebasing from local currencies to the Euro. The jumps in UOM PO nowcasts in Germany in vintages between 2004 and 2005 are the consequence of okun coefficients very close to zero: as discussed, this is a joint problem of the method which cannot handle small okun coefficients and the small sample size which leads to implausibly small coefficient estimates. However, these results would mean very high revisions for the quarters that were first observed in those editions. Since, of course, we are interested in the revision behaviour of the method under the condition that it yields plausible and applicable estimates, we drop all vintages before January 2005 for the following revision analysis.

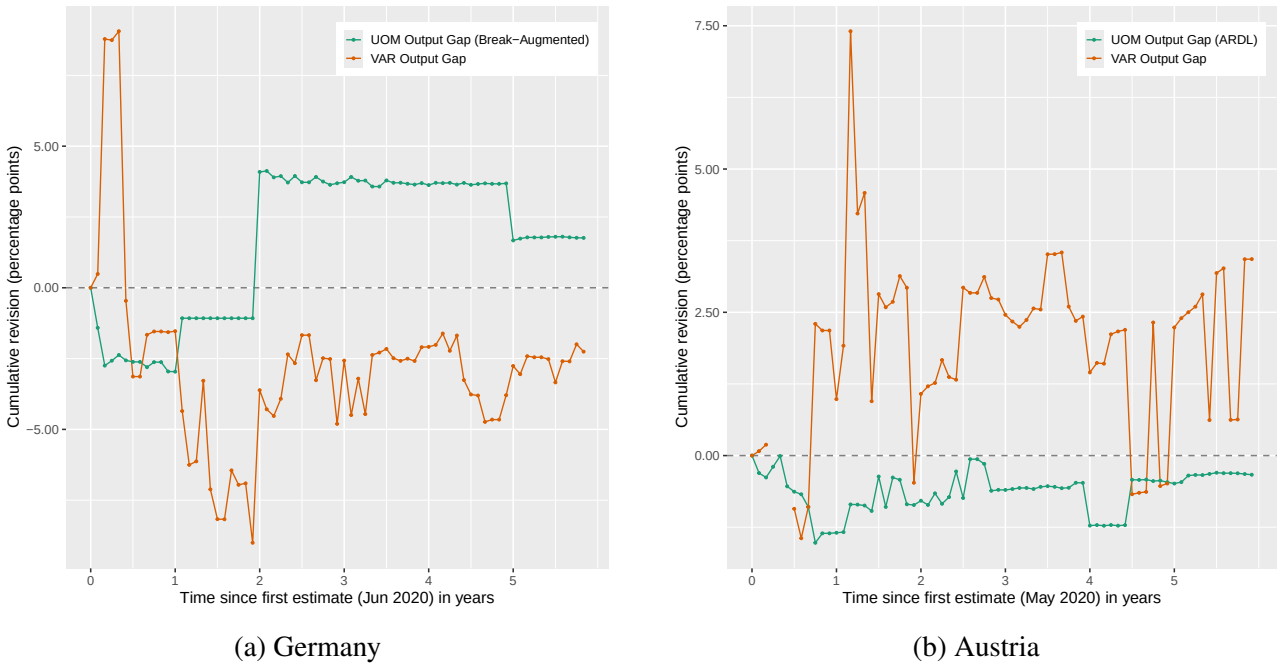


Figure 5: 2020 Q1 Revision over time, in nominal OG changes (percentage points of PO)

Based on this, we can consider the revision behaviour of the methods in single quarters, as shown in Figure 5 the Covid crisis quarter 2020 Q1. Figure 6 shows the average mean revision for the UOM and VAR in Austria and Germany. The mean absolute revision give a picture of the average revision magnitude, no matter the direction of the revision. We can see that both methods have significantly higher average absolute revisions in Germany. This is not only due to the estimation problems of

the UOM, since it also holds true for the VAR. For Austria, both methods have revisions in similar magnitudes and revise the output gap by around 1.5-2% after 2 years, which stays constant for the next 10 years. In Germany, the mean absolute revision of the VAR rises with the time since the first estimate. 10 years after the first estimate, the average revision is around 4%. The German UOM estimation has big revisions, up to 15% 10 years after the first estimate. However, those numbers need to be interpreted with extreme caution because of the discussed implausible PO results.

A very interesting finding is the behaviour of the mean revision in figure 6. We can observe that both methods in both countries have negatively biased revisions after 5 or more years. When assuming that the revisions are more precise estimations than the original estimate, which seems plausible under the condition that one trusts the methodology at all, this would mean that the estimation imposes a positive bias on the OG, which means a restrictive bias on monetary and fiscal policy. While revisions of some magnitude are to be expected, biased revisions are grounds for concern.

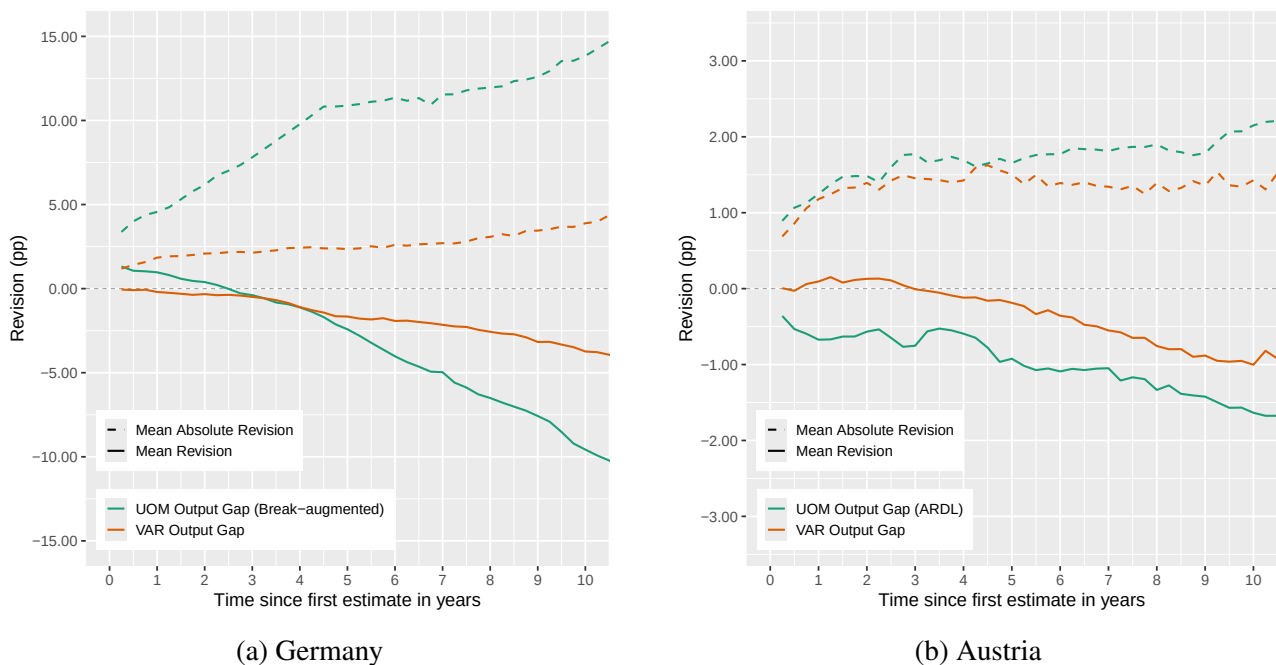
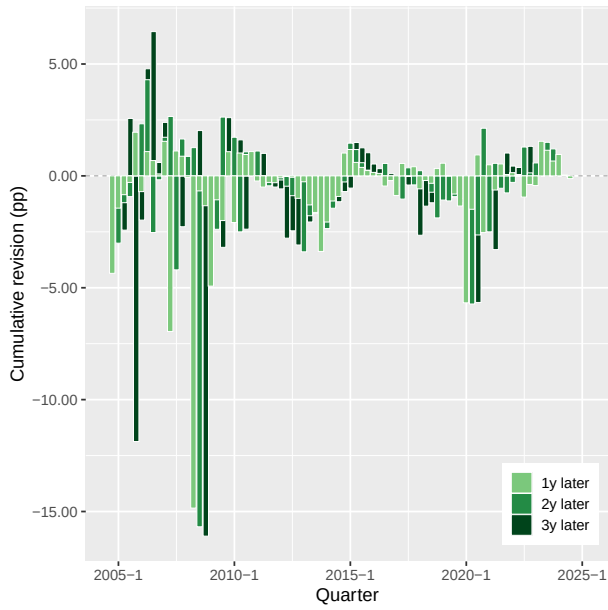
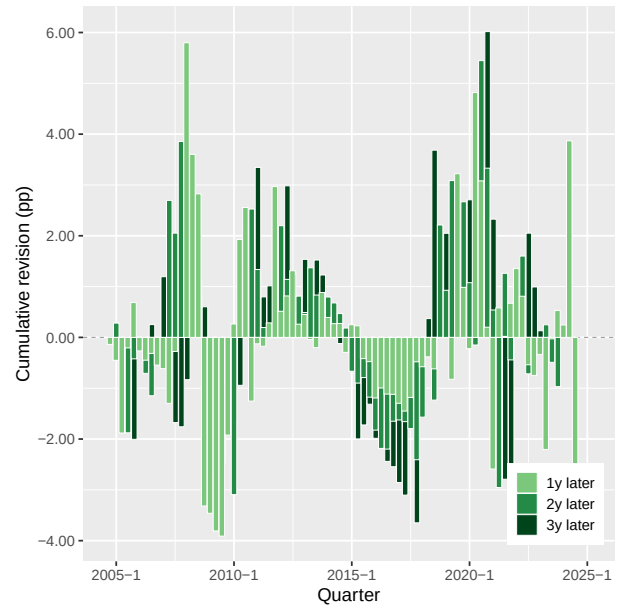


Figure 6: Mean Absolute and Mean Revision over all vintages starting Jan 2005

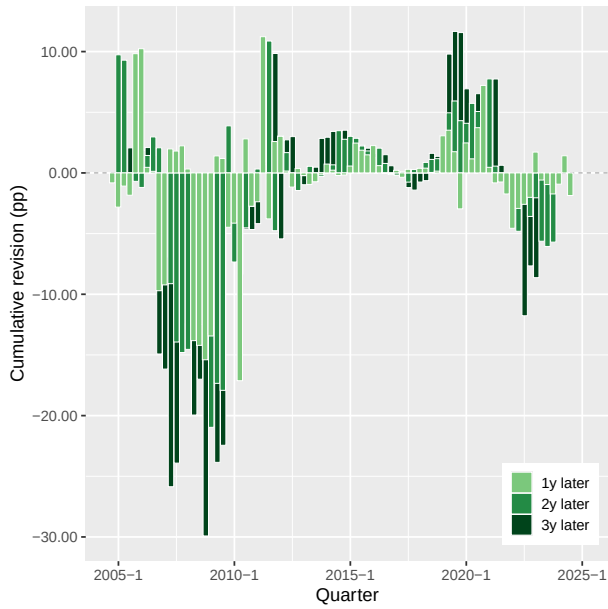
Figure 7 shows the revisions over different horizons dependent on the time period. Particularly interesting are the magnitudes of revisions in crises, which for all measures, are significant. In the GFC, the UOM OG is revised downwards strongly in both Germany and Austria by more than 15% of PO. During the Covid crisis, there are still downwards revisions in Austria, while in Germany, the UOM OG is revised upwards. As stated, the German UOM results, however, might not have a strong meaning. The VAR model also exhibits particularly high revisions in the GFC, up to 15% in Germany and 5% in Austria. Similarly to the UOM, revisions for the covid crisis are still negative but smaller than for the GFC in both countries.



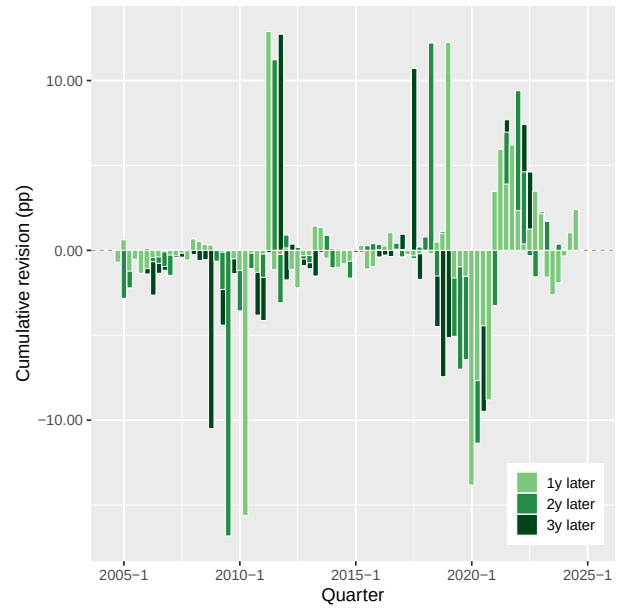
(a) UOM (ARDL), Austria



(b) VAR, Austria



(c) UOM (Break-augmented ARDL), Germany



(d) VAR, Germany

Figure 7: Output Gap revisions by horizon over all analysed vintages, in nominal output gap change (percentage points of PO)

The strong revision behaviour in crises is typical for PO measures, both driven by method problems with handling sudden changes in the data and stronger data revisions in crises (Bierbaumer and Bilek-Steindl 2025). They however also hint towards a fundamental problem of PO estimation measures in general, which is that they tend to be the least reliable when they are needed most: in guiding crises scenarios.

The negative revisions in crises lead to the suspicion of pro-cyclicality. While a broad range of tests required to measure pro-cyclicality in detail would go beyond the scope of this work, we do apply a simple measure in considering the correlation between OG revisions after x years and de-measured GDP growth. The correlation coefficients as well as Spearman ρ s, which we apply in order to not

overinterpret the big outliers that are present in the sample are depicted in Table 3. We can observe statistically significant positive correlation, i.e. signs of pro-cyclicality for the UOM in Austria by means of Spearman ρ over every horizon as well as for the VAR in Germany by means of correlation coefficients. The other values are not statistically significant.

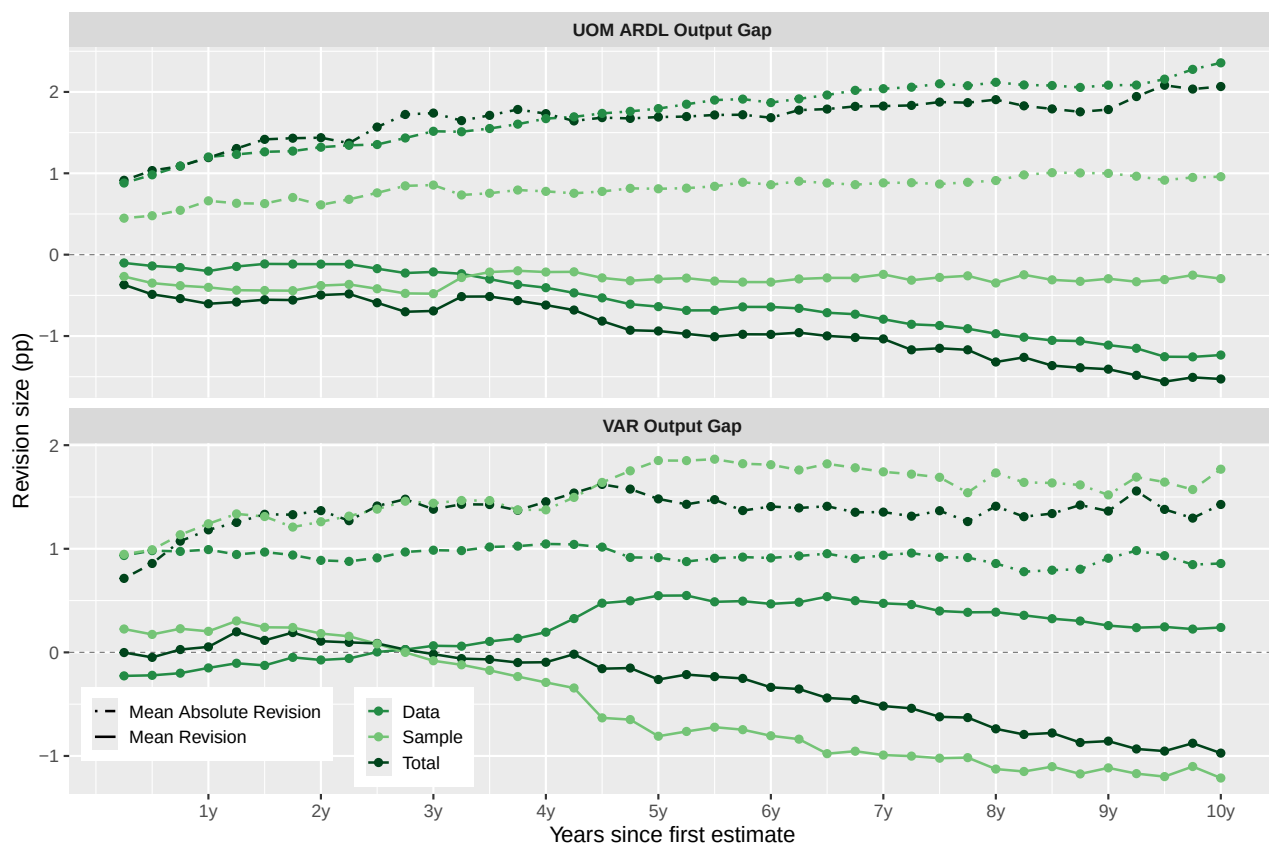
Table 3: Correlation between Output Gap Revisions and Demeaned GDP Growth

Country	Measure	n	1 Year Later		2 Years Later		3 Years Later	
			r	ρ	r	ρ	r	ρ
Austria	UOM (ARDL)	80/76/72	0.182	0.194*	0.146	0.194*	0.151	0.312***
Austria	VAR	78/74/70	0.114	-0.191*	0.146	-0.042	0.111	-0.089
Germany	UOM (Break-ARDL)	80/76/72	0.118	0.163	0.058	0.063	0.038	-0.055
Germany	VAR	80/76/72	0.206*	-0.102	0.245**	0.043	0.305***	0.183

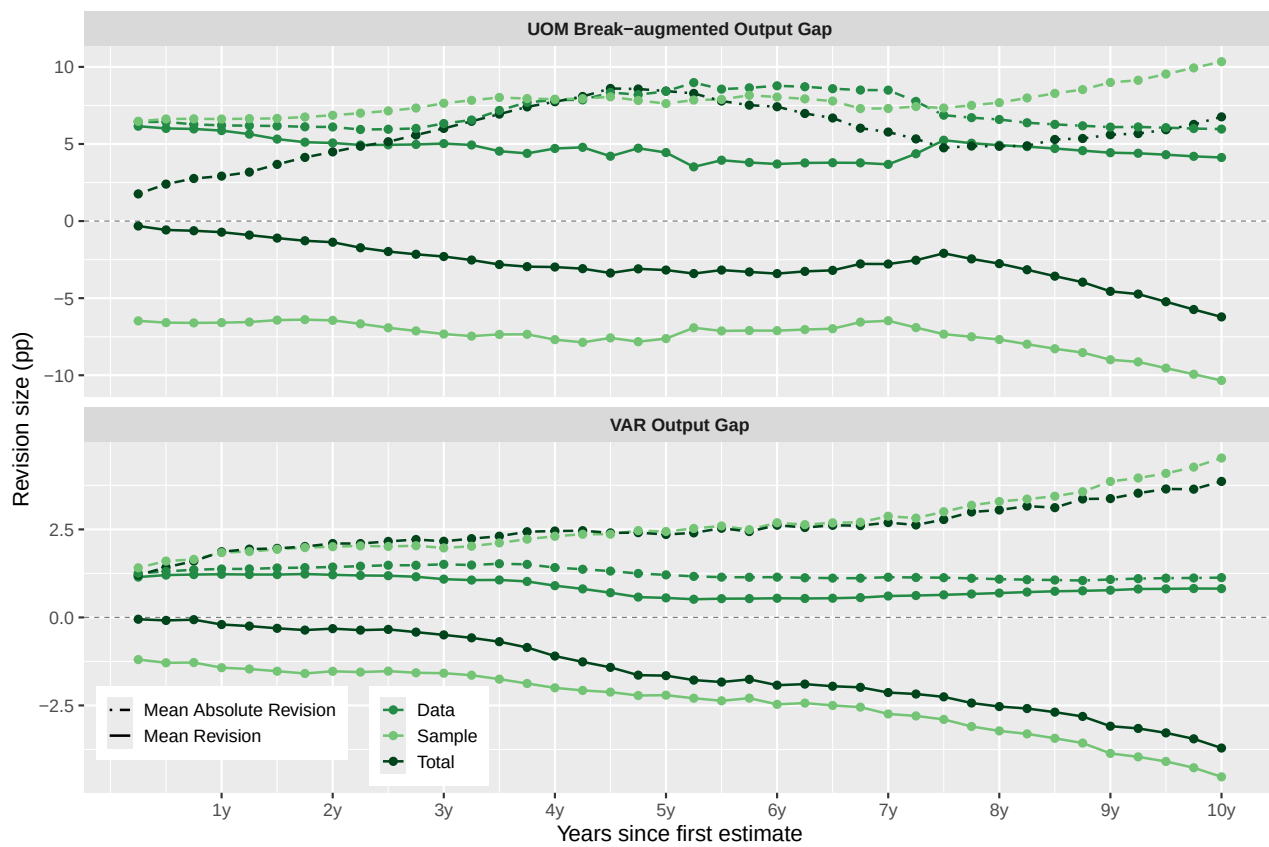
Notes: r = Pearson correlation coefficient; ρ = Spearman rank correlation coefficient. Horizons refer to the revision distance from the first available vintage. GDP growth is demeaned and taken from the first available vintage per quarter. Sample starts 2005 Q1. Significance: ** $p < 0.05$; *** $p < 0.01$.

Finally, we propose a decomposition method in order to derive the difference between the part of the revisions driven by data-revisions and the part driven by data-expansion. To do so, we introduce a counterfactual that re-estimates PO the with the data taken from the point x vintages after the regression, truncated to the length of the dataset at the original first observation. This methodology asks the question: *What if the data revision happened, but the sample expansions did not?* It can therefore isolate the data revision component and interprets the residual as the method expansion component. One caveat is the possible difference in lag structure between the two first estimation and the estimation in a later vintage, since they are individually decided by the AIC. My approach is to estimate the counterfactual on the lag structure of the original estimation in order to get a clean data revision effect. Consequentially, the residual is a combination of the effect of the sample expansion and the effect of the possibly changed lag structure. A further, more grave problem with this methodology is the fact that both the UOM (in case of low okun coefficients) as well as the VAR (in the case of exploding models) tend to give unrealistic results in some vintages, which could as well happen in the newly estimated counterfactuals. Since this would lead to implausibly high or low results of data revisions, and, since they are computed as the residual to a low total revision, as an opposite equally implausible high or low result for sample revision, the counterfactuals that estimate data or sample revisions higher than 40% or lower than -40% are not calculated into the average.

Following Orphanides and Norden (2002), we would expect data revisions to be not the main reason for output gap revisions but still visible and central. Figure 8 shows the decomposition of the revision average as well as the mean absolute revision.



(a) Austria



(b) Germany

Figure 8: Revision average decomposition into sample and data revisions

Considering the mean absolute revision, we can differentiate the relative strength of data revisions and non-data revisions, which we here frame as sample revisions. For the VAR, the main cause of revisions are non-data driven, although data revisions play a relevant role in both countries. For the UOM it is less clear, but data revisions seem to be the stronger cause of revisions, especially in Austria. It is important to remember that our choices of method adjustment to the data were quite prone to induce sample revisions. Including a fixed-lags VAR or a training average of okun coefficients might very well reduce both sample and total revisions. Considering the mean revision can give an explanation where the negative bias in total revisions is coming from. Here it is very notable that while data revisions are either positively or negatively biased depending on method and country, sample revisions are consistently negative through all observations (besides early revisions of the VAR in Austria. This shows that the negative total bias, that we could observe for both methods, cannot be explained away by data revisions.

4.4 Evaluations

While not being framed that way, the Blanchard-Quah VAR is essentially a measure for full-capacity PO. There is no connection to inflation besides the assumption that a full-capacity PO level will coincide with an inflation-stable level. The proposal of Coibion, Gorodnichenko, et al. (2017) to introduce inflation variables into the VAR might lead to further theoretical complications since, without the assumption that inflation-stable PO and full-capacity PO are at the same level, a measure that attempts to estimate both might end up estimating nothing meaningful. Further, the restricting assumption, that permanent shocks have no impact on employment, is a fundamental problem and lacks justification. It cannot be justified by classifying long run shocks as supply shocks, since as we have seen there is the possibility of long run effects of aggregate demand. If some demand effects are included in the permanent shocks, which would simply be a definition change and not a method change, the method still clearly divides into temporary and permanent shocks, arguing that some temporary shocks definitely do not have permanent effects, which can be questioned from an endogeneity perspective. Practically, the method is reliant on the arbitrary choice of an anchor period and with that, a high level of personal judgment. In the frame of this small scale revision analysis, it further shows high negatively biased revisions and partially pro-cyclical, suggesting that it structurally underestimates PO.

For the UOM, conceptually, Fontanari et al. (2021a) are clear that their measure is “entirely unrelated to inflation”, following their scepticism towards any univocal unemployment-inflation relationship. They therefore reject Okuns initial concept of the target as aiming at a high activity but inflation-stable level. While this clearly means that inflation-stable PO cannot be the target of the measure, it also seems implausible that the measure aims for full-capacity PO, since the method does not contain any measure of economic slack besides the distance of the arbitrarily picked unemployment target to the actual unemployment rate. Besides conceptual questions, both methods to pick the target rate of unemployment seem to yield implausible results: The choice of taking the lowest unemployment rate over the sample makes output above PO impossible, which seems questionable no matter the PO concept. Choosing a constant target seems similarly implausible when facing the probable endogeneity of any PO. A further possible issue is the interpretation of the estimated okun relationship. The

fact that these two variables exhibit this relationship does not necessarily mean that a demand push towards an estimated PO would necessarily yield the associated unemployment rate. Further there are different types of demand push, that might affect both output and unemployment very differently and might mean drastic changes in the okun relationship (an extreme example would be a Job Guarantee measure which could drastically reduce unemployment while the demand effect would depend on the paid wage etc.). Further theoretical questions could concern the possibility of endogeneity. Fontanari et al. (2021a, p. 159) state the ability for their method to emit large persistent OGs to be a strength of the method, depicting the perspective that there is no automatic adjustment of output to PO. The possibility of persisting large OGs, however, also demonstrates the absence of a concrete mechanism of PO to adjust to output. If e.g. capital stock adjustment mechanisms or hysteresis effects were present, i.e. a low unemployment would lead to a lower PO, the UOM procedure would be unable to capture those changes unless the target unemployment rate were changed manually. While Fontanari et al. (2020) do recognize hysteresis and calculate a counterfactual long term higher growth path for the case of higher demand, the UOM policy target can not directly be affected by endogeneity effects. In the end, the UOM is probably best understood as the attempt to formulate a simply understandable policy target rather than to empirically pin down a precisely defined PO concept. However, also in regard of the practical implementation, questions remain. The measure is constrained to its identification of a strong okun relationship and cannot produce meaningful results when the okun relationship is small, as we have seen in the Germany exercise. This issue might be overcome when introducing different indicators of labour market slack, as Fontanari et al. (2021a) do. The measure, as estimated here, has significant negatively biased and potentially pro-cyclical revisions and would consequently be hardly fit as a mechanized policy measure. And finally the question remains whether a policy that aims for a high-activity target might be better advised when taking to account a more precise measure of PO rather than a best guess which unemployment rate would be a safe target.

To conclude, in this form, the measures are not consistent enough to be considered as *mechanistic* measures for policy. Besides the theoretical problems, any estimated OG might very well turn out to be revised significantly and even the sign of the OG might switch only a few years later. There is further clear evidence for biased revisions and suggestive evidence for pro-cyclical revisions for the UOM in Austria. When PO is utilized as a policy concept in reality, revisions become an even bigger problem, because the flash estimates necessary to do an actual PO nowcast to rely policy on have quite high revisions as well (Wegmüller et al. 2026).

5 Discussion

There are a few things left to discuss in order to close this work. The following section therefore will go over the implications of the theoretical and empirical work for practical policy implementation as well as limitations, possible critiques and expansions for future research.

5.1 Implications for Fiscal Policy

The purpose of PO estimations is first and foremost to inform policy, in particular fiscal policy. Which implications can we draw from the theoretical and empirical results for PO as a measure to define the volume of fiscal policy? There are two lines of discussion to taken a closer look at here.

Inflation-stable PO or full-capacity PO?

First, we come back to the distinction between full-capacity and inflation-stable output. If we consider PO to be a sensible policy target at all, the first question is, whether inflation-stable PO or full-capacity PO should be the one to aim for for fiscal policy, assuming that it is possible for them to diverge. Intuitively, it seems clear to aim for the inflation-stable level. Since extra demand in an overutilized situation is prone to create inflation, it seems improbable to imagine the inflation-stable PO level significantly above the full-capacity level. In the more imaginable case of inflation-stable level below PO, it does not seem sensible for policy to orient on the output level corresponding to full capacity utilization because the expansionary effects of fiscal policy aiming at this measure would create inflation. There are a few points to consider that question this simple stance.

First, the combination of constant shocks on output and multiple channels for both PO to adjust to output and vice versa might make estimating the inflation-stable level empirically highly complex and hard. In combination with the theoretical dissensus about nature of inflation beyond PO (i.e. accelerating or at a higher constant level), using inflation data to pin the inflation-stable PO level could be an impossible task from a practical perspective. Further, as argued before, there might be connections between the inflation-stable and full-capacity level of PO, since excess capacity might absorb inflationary pressure, which could make full-capacity PO a viable alternative if inflation-stable PO is not possible to measure sensibly. A last argument that would question the simple perspective, that fiscal policy should aim for inflation-stable PO, is the probable endogeneity of inflation-stable PO. Going above the inflation-stable level for a short term could therefore mean an adjustment of PO and permanently higher output at the cost of temporary (or, depending on the approach, permanent) inflation. Blanchard (2018, p. 118) e.g. proposes that there is a “a strong case, although not an overwhelming case, to allow US output to exceed potential for some time, so as to reintegrate some of the workers who left the labour force during the last ten years.”

Consequently, proponents for considering full-capacity PO as a sensible target for fiscal policy could argue that this is the only sensibly calculable measure of the two, that the inflationary danger before might be limited anyway because of the excess capacity and that even if inflation were breached, this might even permanently increase output growth through hysteresis and other endogeneity measures. Further, although right now there is no approach that convincingly estimates an endogenous short-term inflation-stable PO, writing off any estimation of the concept as impossible might be a little short-sighted.

Mechanistic PO measures or room for personal judgment?

We have so far discussed the question, which measure to aim for in a very mechanistic nature, assuming implicitly that the fiscal policy volume would strictly follow a model that estimated PO according to one of the definitions. However, there are good arguments to leave some personal judgment to economists when basing fiscal policy space or monetary policy responses on a PO measure. Skott (2012, p. 109) argues, that “the same theoretical model will not fit all economies. Some economies are labour constrained while others have large amounts of hidden unemployment in backward sectors, and we need a range of models to describe and analyse different circumstances”, which is no less true for empirical estimation methods. But also beyond heterogeneity of countries, the central argument for leaving more room for personal judgement in estimating policy space is the central point that the results of this paper reinforce: The fact that there are blatant problems including high, biased and/or pro-cyclical revisions, structural PO-underestimation, sensitivity to small changes in values, inconsistency with inflation data and fundamental theoretical issues, some of which apply to each approach that could be applied as a mechanistic measure. The case against mechanistic PO measures is furthered even stronger when considering the environment in which they have to make their estimates. Not only are data revisions a constant problem, the initial flash estimate data, which policy has to rely on in real time, have significantly higher revision behaviour, which could lead to estimates which are far away from the actual values, even if the method is conceptually sound. Until those problems are overcome, fiscal policy space estimation needs to include personal judgment calls.

An idea on how personal judgment as well as information from econometric estimations of full-capacity and/or inflation-stable PO can be integrated into a policy framework is given in Hein (2023a), who argues for fiscal policy to occasionally “test the waters”, trying to get as close to the inflation-stable level as possible and fighting or accepting occasional inflation in order to use hysteresis effects, possibly reaping the benefits of higher growth in the mid-term. For the initial estimation of policy reactions to shocks, the UOM could provide a simple policy framework, although its partial conceptual unclarity. A further potential measure to give an estimation of full-capacity PO to consider in the design of fiscal policy space is the measure by Bassi (2024), which however also needs to be evaluated further before being relied on too much. This policy framework also has to consider other limits than the inflation limits to fiscal space expansion, such as Balance-of-Payments constraints (León-Ledesma and Thirlwall 2002) or, of course, ecological limits to growth. Both unquestionably constitute significant constraints for PO expansion in at least big parts of the world (or the entire world in the case of ecological limits to growth). They are, however, in this context, seen as separate constraints, which therefore do not necessarily fall under the constructed concept of PO. However, taking the concept of a maximum reachable output under given constraints seriously might make constructing a PO that includes these constraints an interesting, while enormous, further research agenda.

5.2 Limitations and Avenues for Future Research

One limitation, which is frequently suggested in the literature, is the incapability of the unemployment rate to properly estimate labour market slack. In both Germany and Austria, there are multiple

measures and counting techniques for (long-term) unemployment along, which, at times, differ substantially in magnitude (e.g. see Picek (2020) for Austria). Fontanari et al. (2021a) therefore expand the UOM measure by using different labour underutilization methods. This is both interesting for robustness, but also to understand how the normative definition of common unemployment measures affects PO. Utilizing different labour market slack measures could be helpful in further understanding the relation between the level of output and labour market slack in Germany and possibly lead to more plausible PO results. Taking into consideration the scarcity of vintage data this would require either constructing vintage data of alternative labour market slack measures, assuming that they are not revised over time or constraining the analysis to the current vintage, none of which was possible *and* sensible in the context of this work. It could however be a possible avenue for future research. Further possible expansions could include the implementation of different evaluation criteria for PO measures. A possible proposition is the empirical ability to predict inflation. This however would need to include assumptions about the nature of the inflation-unemployment relationship, which might include problems based on its uncertain nature. For the VAR, the model estimated could be criticized as very simplistic, since it is excluding structural break estimations and the proposed expansions by Coibion, Gorodnichenko, et al. (2017). This is partially for the sake of time and space constraints and partially for scepticism about a measure that tries to estimate inflation-stable PO and full-capacity inflation at the same time. Accounting for structural breaks, however, should be a sensible expansion. A further expansion could be identifying the SVAR through a different econometric approach, such as a narrative approach or sign restrictions, aiming to overcome some of the problematic assumptions and turn it into a strong measure to identify full-capacity PO. Central expansions also could evaluate more recently discussed measures. Especially interesting would be a further investigation of the measure of Bassi (2024), who seems to provide a conceptually clear measure of full-capacity PO. Further, the DFM by Jarociński and Lenza (2018) could be an interesting next method to evaluate. Beyond this, an expansion of the sample size of analysed countries for which there are broad real-time data could significantly help in order to achieve a broader more robust picture of revision behaviour of the single methods. Finally, further research should consider looking into estimations of the difference between inflation-stable PO and full-capacity PO. If they diverge regularly, this could mean detrimental identification problems for any method that tries to evaluate both at the same time.

6 Conclusion

This work set out to evaluate alternatives to the estimation methodologies used in international institutions. On the theoretical side, it introduces a new potential problem of PO estimation methods: conceptual inaccuracy due to assuming inflation-safe PO and full-capacity PO to be the same output level. Since it is unclear how big and regular gaps between those two levels are, it also is unclear whether assuming them to be the same level has grave consequences for PO estimation methods. The revision results as well as the theoretical analysis of the UOM and the BQ-VAR cast strong doubts on whether the methods could sensibly be implemented as mechanistic tools to estimate fiscal policy space. While the search for clearly viable mechanistic replacement methods for the methods em-

ployed in international institutions will have to go on, this should not give the false impression that the status quo of PO and fiscal space estimation in (inter-)national institutions could not be largely improved without an optimal estimation method. An alternative approach should include implementing a “testing-the-waters” approach for fiscal policy, trying to stay around the inflation-safe PO, possibly temporarily going above it to use hysteresis effects. The idea that fiscal policy needs to act far more aggressively in order to avoid negative fiscal policy-PO estimation loops (Heimberger 2020; Fatás 2019) is slowly reaching parts of the economic mainstream. It will remain to be seen whether and how quickly this development will affect the PO implementation in international institutions or whether it will need more global economic crises to become clear.

References

- Arestis, Philip and Malcolm Sawyer (2002). ‘*New Consensus, New Keynesianism, and the Economics of the ‘Third Way’*. Working paper. Levy Economics Institute. URL: https://EconPapers.repec.org/RePEc:lev:wrkpap:wp_364 (visited on 09/26/2026).
- Bai, Jushan and Pierre Perron (1998). “Estimating and Testing Linear Models with Multiple Structural Changes”. In: *Econometrica* 66.1, pp. 47–78. ISSN: 0012-9682. DOI: 10.2307/2998540.
- Ball, Laurence M (2014). “Long-Term Damage from the Great Recession in OECD Countries”. In: *NBER Working Paper* 20185. URL: <https://www.nber.org/papers/w20185>.
- Bassi, Federico (Dec. 1, 2024). “Excess capacity and hysteresis in EU Countries. A structural approach”. In: *Structural Change and Economic Dynamics* 71, pp. 116–134. ISSN: 0954-349X. DOI: 10.1016/j.strueco.2024.07.002.
- Bierbaumer, Jürgen and Sandra Bilek-Steindl (2025). “Die WIFO-Schnellschätzung: Revisionsanalyse und Kurzbeschreibung der Methode”. In: *WIFO Monatsberichte (monthly reports)* 98.12, pp. 667–679. URL: <https://ideas.repec.org/a/wfo/monber/y2025i12p667-679.html>.
- Blanchard, Olivier (2018). “Should We Reject the Natural Rate Hypothesis?” In: *Journal of Economic Perspectives* 32.1, pp. 97–120. ISSN: 0895-3309. DOI: 10.1257/jep.32.1.97.
- Blanchard, Olivier and Danny Quah (1989). “The Dynamic Effects of Aggregate Demand and Supply Disturbances”. In: *National Bureau of Economic Research Working Paper Series* No. 2737. DOI: 10.3386/w2737.
- Blanchard, Olivier and Lawrence H. Summers (1986). “Hysteresis and the European Unemployment Problem”. In: *NBER Macroeconomics Annual* 1, pp. 15–78. ISSN: 0889-3365. DOI: 10.2307/3585159.
- (2017). *Rethinking Stabilization Policy: Evolution or Revolution?* DOI: 10.3386/w24179.
- Carlin, Wendy and David Soskice (1990). *Macroeconomics and the Wage Bargain: A Modern Approach to Employment, Inflation, and the Exchange Rate*. Oxford: Oxford University Press. URL:

<https://www.inet.ox.ac.uk/publications/macroeconomics-and-the-wage-bargain-a-modern-approach-to-employment-inflation-and-the-exchange-rate>.

Carlin, Wendy and David Soskice (2009). “Teaching intermediate Macroeconomics using the three equation model”. In: *Macroeconomic Theory and Macroeconomic Pedagogy*. London: Palgrave Macmillan UK. ISBN: 978-0-230-29166-9. DOI: 10.1007/978-0-230-29166-9_1.

CBO (2014). *Revisions to CBO’s Projections of Potential Output Since 2007*. Congressional Budget Office. URL: <https://www.cbo.gov/sites/default/files/113th-congress-2013-2014/reports/45150-PotentialOutput-OneColumn.pdf> (visited on 05/23/2026).

Cerra, Valerie, Antonio Fatás, et al. (Mar. 2023). “Hysteresis and Business Cycles”. In: *Journal of Economic Literature* 61.1, pp. 181–225. ISSN: 0022-0515. DOI: 10.1257/jel.20211584.

Cerra, Valerie and Sweta Saxena (Mar. 2000). “Alternative Methods of Estimating Potential Output and Output Gap: An Application to Sweden”. In: *IMF Working Paper* 2000/059. URL: <https://ssrn.com/abstract=3923379>.

Claeys, Grégory et al. (2016). *A proposal to revive the European fiscal framework*. Research Report 2016/07. Bruegel Policy Contribution. URL: <https://www.econstor.eu/handle/10419/165974>.

Coibion, Olivier, UT Austin, et al. (2018). *Real-Time Estimates of Potential GDP: Should the Fed Really Be Hitting the Brakes?* Center on Budget and Policy Priorities.

Coibion, Olivier, Yuriy Gorodnichenko, et al. (2017). “The Cyclical Sensitivity in Estimates of Potential Output”. In: Working Paper Series. DOI: 10.3386/w23580.

Dray, Mark and A. P. Thirlwall (2011). “The endogeneity of the natural rate of growth for a selection of Asian countries”. In: *Journal of Post Keynesian Economics* 33.3. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/03098161003677330303>, pp. 451–468. ISSN: 0160-3477. DOI: 10.2753/PKE0160-3477330303.

Dunsch, Sophie (2016). “Okun’s Law and Youth Unemployment in Germany and Poland”. In: *International Journal of Management and Economics* 49.1, pp. 34–57. ISSN: 2299-9701. DOI: 10.1515/ijme-2016-0003.

Dutt, Amitava Krishna (1987). “Alternative closures again: a comment on ‘Growth, distribution and inflation’”. In: *Cambridge Journal of Economics* 11.1, pp. 75–82. ISSN: 0309-166X. URL: <https://www.jstor.org/stable/23598097> (visited on 12/13/2024).

ECB (2026). *Datasets — ECB Data Portal*. URL: <https://data.ecb.europa.eu/data/datasets> (visited on 05/31/2026).

- Edge, Rochelle M. and Jeremy B. Rudd (2016). “Real-Time Properties of the Federal Reserve’s Output Gap”. In: *The Review of Economics and Statistics* 98.4, pp. 785–791. ISSN: 0034-6535. DOI: 10.1162/REST_a_00555.
- Eichner, Alfred S. (1976). *The megacorp and oligopoly: micro foundations of macro dynamics*. Cambridge [Eng.] ; New York: Cambridge University Press. 365 pp. ISBN: 978-0-521-20885-7.
- Fatás, Antonio (2019). “Fiscal Policy, Potential Output, and the Shifting Goalposts”. In: *IMF Economic Review* 67.3, pp. 684–702. ISSN: 2041-417X. DOI: 10.1057/s41308-019-00087-z.
- Federal Reserve Bank of Dallas et al. (2011). “A Real-Time Historical Database for the OECD”. In: *Federal Reserve Bank of Dallas, Globalization and Monetary Policy Institute Working Papers* 2011.96. DOI: 10.24149/gwp96.
- Felipe, J. and J.S.L. McCombie (2014). “The Aggregate Production Function: ‘Not Even Wrong’”. In: *Review of Political Economy* 26.1. eprint: <https://doi.org/10.1080/09538259.2013.874192>, pp. 60–84. ISSN: 0953-8259. DOI: 10.1080/09538259.2013.874192.
- Fellinger, Serafin (2022). *Austria’s potential output*. Universität Wien. DOI: 10.25365/thesis.71432.
- Fleischman, Charles A and John M Roberts (2011). “Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.” In.
- Fontanari, Claudia et al. (2020). “Potential Output in Theory and Practice: A Revision and Update of Okun’s Original Method”. In: *Structural Change and Economic Dynamics* 54, pp. 247–266. ISSN: 0954-349X. DOI: 10.1016/j.strueco.2020.04.008.
- (2021a). “The updated Okun method for estimation of potential output with alternative measures of labor underutilization”. In: *Structural Change and Economic Dynamics* 60, pp. 158–178. ISSN: 0954-349X. DOI: 10.1016/j.strueco.2021.11.008.
 - (2021b). *The Updated Okun Method for Estimation of Potential Output with Broad Measures of Labor Underutilization: An Empirical Analysis*. Rochester, NY. URL: <https://papers.ssrn.com/abstract=3919291> (visited on 11/08/2025).
- Foroni, Claudia et al. (2018). “Using low frequency information for predicting high frequency variables”. In: *International Journal of Forecasting* 34.4, pp. 774–787. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2018.06.004.
- Giannone, Domenico et al. (2012). “An Area-Wide Real-Time Database for the Euro Area”. In: *The Review of Economics and Statistics* 94.4, pp. 1000–1013. ISSN: 0034-6535. DOI: 10.1162/REST_a_00317.

- Havik, Karel et al. (2014). “The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps”. In: *European Economy - Economic Papers 2008 - 2015*. Number: 535. URL: <https://ideas.repec.org/p/euf/ecopap/0535.html> (visited on 02/16/2026).
- Heimberger, Philipp (2020). “Potential Output, EU Fiscal Surveillance and the COVID-19 Shock”. In: *Intereconomics* 55.3, pp. 167–174. ISSN: 1613-964X. DOI: 10.1007/s10272-020-0895-z.
- Heimberger, Philipp and Jakob Kapeller (2017). “The performativity of potential output: pro-cyclicality and path dependency in coordinating European fiscal policies”. In: *Review of International Political Economy* 24.5. eprint: <https://doi.org/10.1080/09692290.2017.1363797>, pp. 904–928. ISSN: 0969-2290. DOI: 10.1080/09692290.2017.1363797.
- Heimberger, Philipp, Jakob Kapeller, and Bernhard Schütz (2017). “The NAIRU determinants: What’s structural about unemployment in Europe?” In: *Journal of Policy Modeling* 39.5, pp. 883–908. ISSN: 0161-8938. DOI: 10.1016/j.jpolmod.2017.04.003.
- Hein, Eckhard (2004). “Die NAIRU – eine post-keynesianische Interpretation”. In: *European Journal of Economics and Economic Policies* 1.1, pp. 43–66. ISSN: 1613-0960, 2195-3376. DOI: 10.4337/ejeep.2004.01.07.
- (2023a). “A post-Keynesian co-ordinated macroeconomic policy mix”. In: *Macroeconomics after Kalecki and Keynes*. Section: Macroeconomics after Kalecki and Keynes. Edward Elgar Publishing, pp. 197–215. ISBN: 978-1-80392-728-2. URL: <https://www.elgaronline.com/monochap/book/9781803927282/book-part-9781803927282-11.xml> (visited on 05/29/2026).
 - (Jan. 17, 2023b). “Post-Keynesian macroeconomic models with conflict inflation”. In: *Macroeconomics after Kalecki and Keynes*. Section: Macroeconomics after Kalecki and Keynes. Edward Elgar Publishing, pp. 136–196. ISBN: 978-1-80392-728-2. URL: <https://www.elgaronline.com/monochap/book/9781803927282/book-part-9781803927282-10.xml> (visited on 05/29/2026).
 - (2024). “Inflation is always and everywhere ... a conflict phenomenon: post-Keynesian inflation theory and energy price driven conflict inflation, distribution, demand and employment”. In: *European Journal of Economics and Economic Policies* 21.2, pp. 202–231. ISSN: 2052-7764, 2052-7772. DOI: 10.4337/ejeep.2024.0135.
- Hein, Eckhard and Artur Tarassow (2010). “Distribution, aggregate demand and productivity growth: theory and empirical results for six OECD countries based on a post-Kaleckian model”. In: *Cambridge Journal of Economics* 34.4, pp. 727–754. ISSN: 0309-166X. DOI: 10.1093/cje/bep066.
- Horn, Gustav et al. (2007). *Methodische Fragen mittelfristiger gesamtwirtschaftlicher Projektionen am Beispiel des Produktionspotenzials*. Research Report 1/2007. IMK Study. URL: <https://www.econstor.eu/handle/10419/106217> (visited on 02/16/2026).

- Jarociński, Marek and Michele Lenza (2018). “An Inflation-Predicting Measure of the Output Gap in the Euro Area”. In: *Journal of Money, Credit and Banking* 50.6. eprint: <https://onlinelibrary.wiley.com/doi/pdf/pp.1189-1224>. ISSN: 1538-4616. DOI: 10.1111/jmcb.12496.
- Krugman, Paul (2018). “Opinion — The Economic Future Isn’t What It Used to Be (Wonkish)”. In: *The New York Times*. ISSN: 0362-4331. URL: <https://www.nytimes.com/2018/09/30/opinion/the-economic-future-isnt-what-it-used-to-be-wonkish.html> (visited on 09/26/2026).
- Lanzafame, Matteo (2010). “The endogeneity of the natural rate of growth in the regions of Italy”. In: *International Review of Applied Economics* 24.5. eprint: <https://doi.org/10.1080/02692170903426039>, pp. 533–552. ISSN: 0269-2171. DOI: 10.1080/02692170903426039.
- Lavoie, Marc (2022). *Post-Keynesian Economics: New Foundations*. Edward Elgar Publishing. ISBN: 978-1-83910-962-1 978-1-83910-961-4. DOI: 10.4337/9781839109621.
- (2024). “Conflictual Inflation and the Phillips Curve”. In: *Review of Political Economy* 36.4. eprint: <https://doi.org/10.1080/09538259.2023.2294305>, pp. 1397–1419. ISSN: 0953-8259. DOI: 10.1080/09538259.2023.2294305.
- León-Ledesma, Miguel A. and A. P. Thirlwall (2002). “The endogeneity of the natural rate of growth”. In: *Cambridge Journal of Economics* 26.4, pp. 441–459. ISSN: 0309-166X. URL: <https://www.jstor.org/stable/23602024> (visited on 05/18/2026).
- Mishkin, Frederic S. (2007). “Estimating potential output: a speech at the Conference on Price Measurement for Monetary Policy, Federal Reserve Bank of Dallas, Dallas, Texas, May 24, 2007”. In: *Speech*. Number: 299. URL: <https://ideas.repec.org/p/fip/fedgsq/299.html> (visited on 11/30/2025).
- Nebot, César et al. (2019). “New insights into the nonlinearity of Okun’s law”. In: *Economic Modelling* 82, pp. 202–210. ISSN: 0264-9993. DOI: 10.1016/j.econmod.2019.01.005.
- OECD (2026). *OECD Data Explorer • Short-term economic statistics revisions*. URL: https://data-explorer.oecd.org/vis?lc=en&fs%5B0%5D=Topic%2C1%7CEconomy%23ECO%23%7CShort-term%20economic%20statistics%23ECO_STS%23&pg=0&fc=Topic&bp=true&snb=20&vw=tb&df%5Bds%5D=dsDisseminateFinalDMZ&df%5Bid%5D=DSD_STES_REVISIONS%40DF_STES_REVISIONS&df%5Bag%5D=OECD.SDD.STES&df%5Bvs%5D=4.0&pd=%2C&dq=FRA.M....&ly%5Brw%5D=TIME_PERIOD&ly%5Bcl%5D=EDITION&ly%5Brs%5D=MEASURE&to%5BTIME_PERIOD%5D=false (visited on 05/08/2026).
- Okun, Arthur M (1962). *Potential GNP: its measurement and significance*. Cowles Foundation for Research in Economics at Yale University.

- Orphanides, Athanasios and Simon van Norden (2002). “The Unreliability of Output-Gap Estimates in Real Time”. In: *The Review of Economics and Statistics* 84.4, pp. 569–583. ISSN: 0034-6535. URL: <https://www.jstor.org/stable/3211719> (visited on 02/09/2026).
- Palumbo, Antonella (2015). “Studying Growth in the Modern Classical Approach: Theoretical and Empirical Implications for the Analysis of Potential Output”. In: *Review of Political Economy* 27.3. eprint: <https://doi.org/10.1080/09538259.2015.1067366>, pp. 282–307. ISSN: 0953-8259. DOI: 10.1080/09538259.2015.1067366.
- Picek, Oliver (2020). “Eine Jobgarantie für Österreichs Langzeitarbeitslose”. In: *Momentum Quarterly - Zeitschrift für sozialen Fortschritt* 9.2, pp. 103–126. ISSN: 2226-5538. DOI: 10.15203/MOMENTUMQUARTERLY.VOL9.NO2.P103-126.
- Primavesi, Leo (2024). “The (dys)functionality of the aggregate production function”. In: *ZÖSS Hamburg. ExMa-Papers*. ISSN: 1868-5005/56. URL: https://savearchive.zbw.eu/bitstream/11159/653714/1/1886191042_0.pdf.
- (2026). *Replication Package for this Work*. URL: <https://gitlab.com/LeoPrimavesi/replication-package-master-thesis> (visited on 05/31/2026).
- Proietti, Tommaso et al. (2020). “A Systemic Approach to Estimating the Output Gap for the Italian Economy”. In: *Comparative Economic Studies* 62.3, pp. 465–493. ISSN: 1478-3320. DOI: 10.1057/s41294-020-00127-y.
- Resende, Carlos de (2014). *An Assessment of IMF Medium-Term Forecasts of GDP Growth*. BP/14/01. Independent Evaluation Office, IMF.
- Rowthorn, R. E. (1977). “Conflict, inflation and money”. In: *Cambridge Journal of Economics* 1.3, pp. 215–239. ISSN: 0309-166X. DOI: 10.1093/oxfordjournals.cje.a035360.
- Skott, Peter (2012). “Theoretical and Empirical Shortcomings of the Kaleckian Investment Function”. In: *Metroeconomica* 63.1. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-999X.2010.04111.x>, pp. 109–138. ISSN: 1467-999X. DOI: 10.1111/j.1467-999X.2010.04111.x.
- Steindl, Josef (1952). *Maturity and stagnation in American capitalism*. In collab. with Internet Archive. Oxford : Blackwell. 272 pp. URL: <http://archive.org/details/maturitystagnati0000stei> (visited on 02/09/2026).
- Topel, Robert (1990). “Specific capital and unemployment: Measuring the costs and consequences of job loss”. In: *Carnegie-Rochester Conference Series on Public Policy* 33, pp. 181–214. ISSN: 0167-2231. DOI: 10.1016/0167-2231(90)90013-B.
- Vogel, Lena (2009). “The endogeneity of the natural rate of growth – an empirical study for Latin-American countries”. In: *International Review of Applied Economics* 23.1. eprint: <https://doi.org/10.1080/02692170802496869>, pp. 41–53. ISSN: 0269-2171. DOI: 10.1080/02692170802496869.

Wang, Qingwei et al. (2009). *Projecting Potential Output: Methods and Problems*. Vol. 42. ZEW Economic Studies. Heidelberg: Physica-Verlag HD. ISBN: 978-3-7908-2176-5 978-3-7908-2177-2. DOI: 10.1007/978-3-7908-2177-2.

Wegmüller, Philipp et al. (2026). “GDP-Flash Estimates: An International Assessment”. In: URL: <https://crawford.anu.edu.au/cama/research/gdp-flash-estimates-international-assessment>.

Appendix

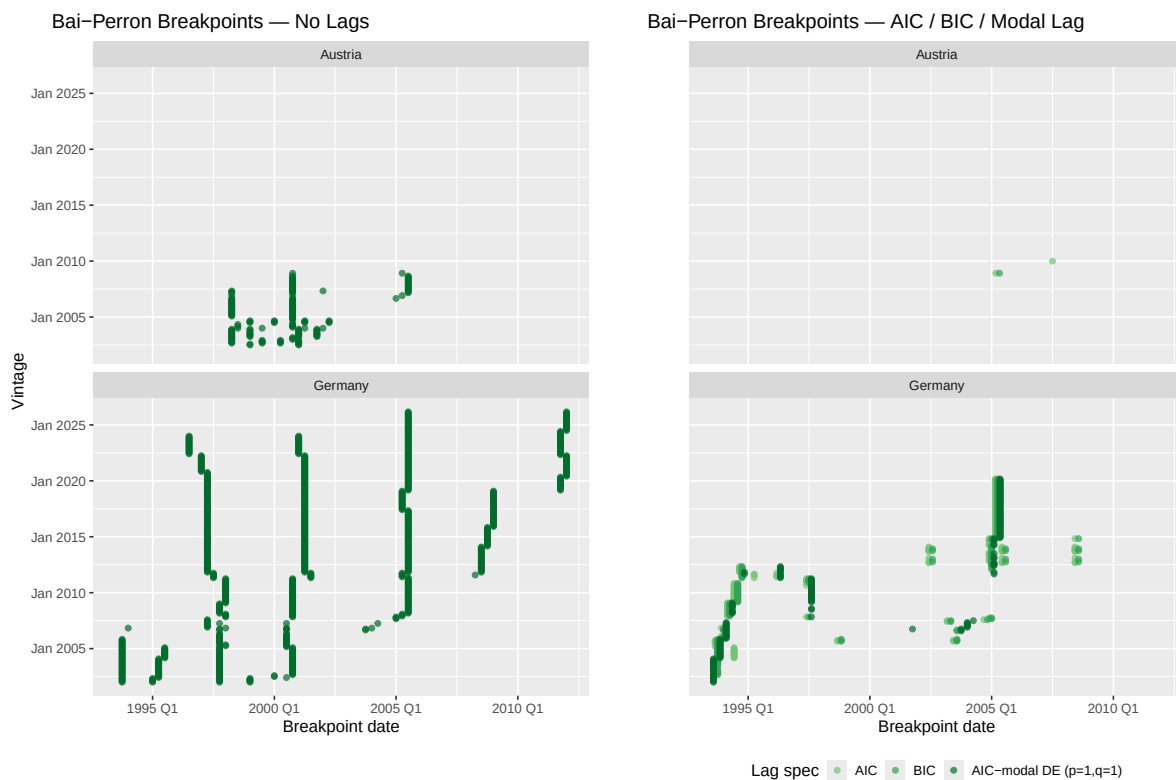
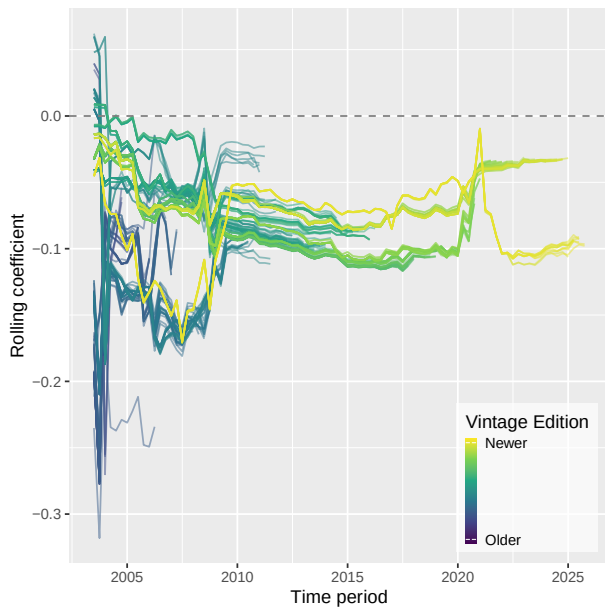
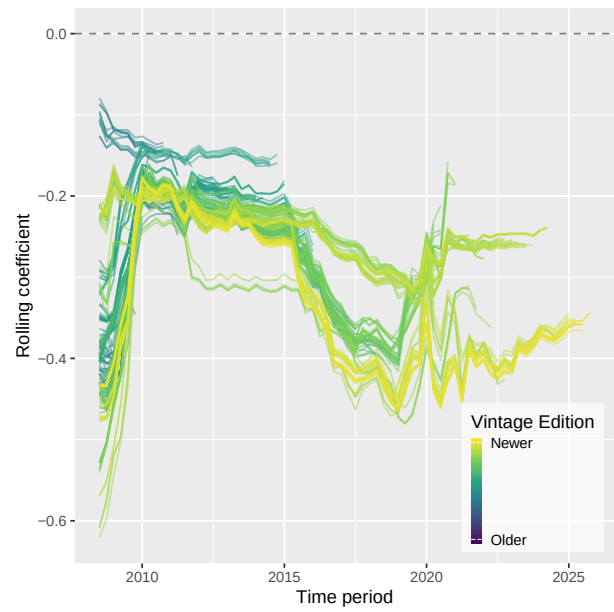


Figure A1: Breakpoints with different assumed regression structures across vintages



(a) Germany



(b) Austria

Figure A2: 50 quarter rolling okun coefficient timelines over different vintages.

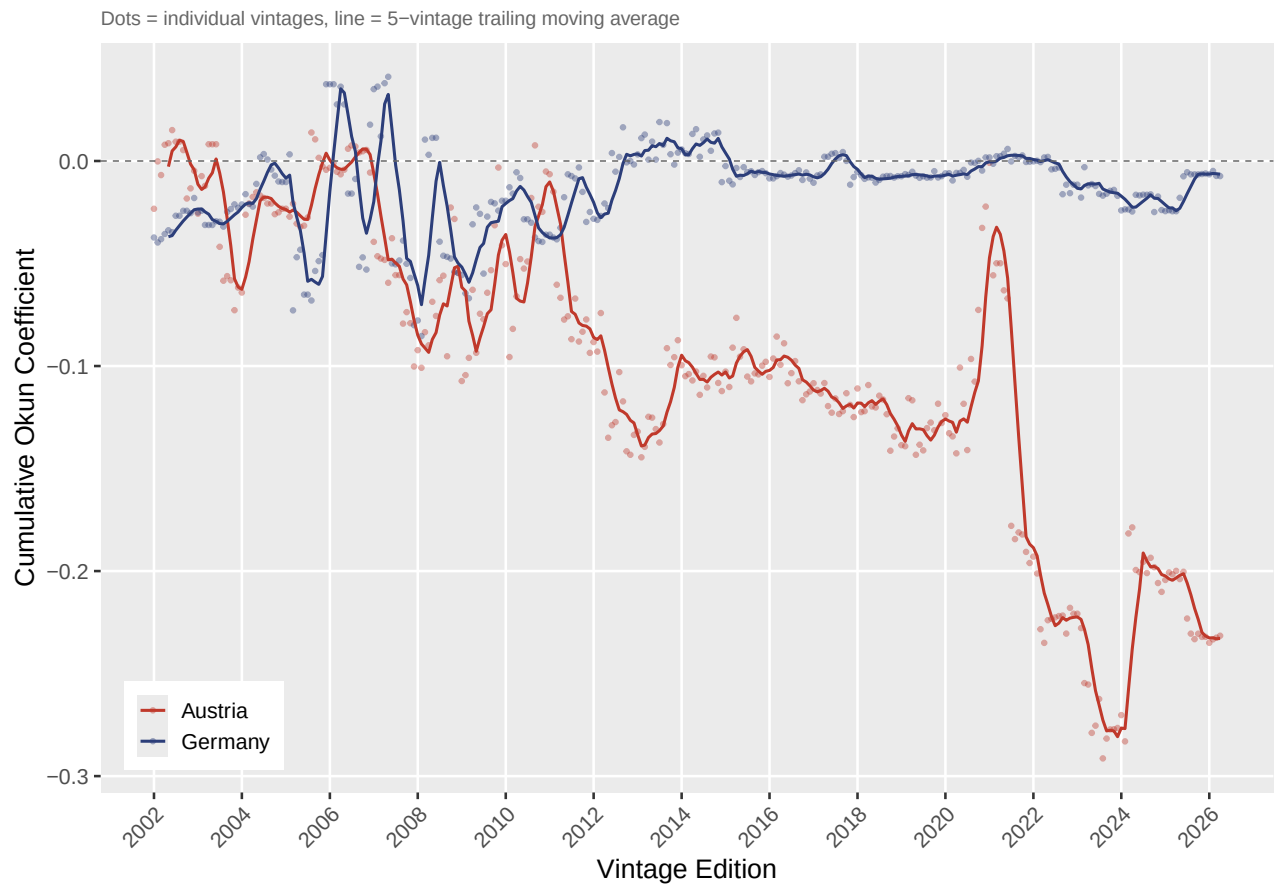


Figure A3: Cumulative Okun Coefficients estimated by RU-MIDAS-Model

Table A1: Share of Explosive VAR Models Across Vintages by Lag Specification

Lag Method	Specification	Austria (AT)	Germany (DE)
Fixed lags	$p = 2$	2.4%	0.0%
	$p = 3$	3.7%	0.0%
	$p = 4$	4.7%	0.0%
	$p = 8$	25.0%	1.7%
AIC	$p_{\max} = 2$	2.4%	0.0%
	$p_{\max} = 3$	3.7%	0.0%
	$p_{\max} = 4$	4.0%	0.0%
	$p_{\max} = 5$	6.7%	0.0%
	$p_{\max} = 6$	10.8%	0.0%
	$p_{\max} = 7$	32.0%	0.0%
	$p_{\max} = 8$	23.0%	0.0%

Notes: Each cell reports the share of explosive VAR models (i.e., with at least one eigenvalue outside the unit circle) across all available vintages for the given lag specification. The preferred specification (comparably low explosion rate while high lags) is **AIC with $p_{\max} = 4$** for Austria, and **AIC with $p_{\max} = 8$** for Germany.