

EFZG WORKING PAPER SERIES  
EFZG SERIJA ČLANAKA U NASTAJANJU  
ISSN 1849-6857  
UDK 33:65

Br. 26-06

Josip Tica, Vladimir Arčabić i Viktor Viljevac

## **Output Gap and Model Uncertainty**



SVEUČILIŠTE U  
ZAGREBU



## Output Gap and Model Uncertainty

Josip Tica  
jtica@efzg.hr  
Ekonomski fakultet  
Sveučilište u Zagrebu  
Trg J. F. Kennedy 6  
10 000 Zagreb, Croatia

Izv.prof. dr. sc. Vladimir Arčabić  
varcabic@efzg.hr  
Ekonomski fakultet  
Sveučilište u Zagrebu  
Trg J. F. Kennedy 6, 10000 Zagreb, Hrvatska  
i  
Hrvatska akademija znanosti i umjetnosti  
Trg Nikole Šubića Zrinskog 11, 10000 Zagreb  
Član suradnik

Viktor Viljevac  
vviljevac@efzg.hr  
Ekonomski fakultet  
Sveučilište u Zagrebu  
Trg J. F. Kennedy 6  
10 000 Zagreb, Croatia

Stajališta iznesena u ovom članku u nastajanju stavovi su autora te ne predstavljaju stavove Ekonomskog fakulteta Zagreb. Članak nije prošao formalnu recenziju i odobrenje. Članak je objavljen kako bi dobio komentare o istraživanjima u tijeku, prije nego što se pojavi u konačnom obliku u akademskom časopisu ili na nekom drugom mjestu.

Copyright February 2026 by Josip Tica, Vladimir Arčabić and Viktor Viljevac

Sva prava pridržana.

Dijelovi teksta mogu biti navedeni pod uvjetom da se u potpunosti navede izvor.

# Output Gap and Model Uncertainty\*

Josip Tica<sup>†a</sup>, Vladimir Arčabić<sup>‡a, b, c</sup>, and Viktor Viljevac<sup>§a</sup>

<sup>a</sup>Faculty of Economics and Business, University of Zagreb  
Trg J.F.Kennedyja 6 10000, Zagreb, Croatia

<sup>b</sup>The Croatian Academy of Sciences and Arts

Trg Nikole Šubića Zrinskog 11, 10000 Zagreb, Croatia

<sup>c</sup>Institute for Development and International Relations  
Lj. F. Vukotinovića 2, 10000 Zagreb, Croatia

September 2026

## Abstract

Output gap estimates play a key role in macroeconomic policy, but different estimation methods often give conflicting readings of the cyclical position. This paper explores the quasi-real-time dynamics of model uncertainty, defined as the cross-method dispersion of output gap estimates. To estimate output gaps and quantify model uncertainty, we use thirteen univariate, multivariate, and institutional output gap measures for 25 EU member states and the United Kingdom over 2002:Q1 to 2024:Q4. The results show an asymmetric pattern: model uncertainty rises sharply in contractions but shows no clear relationship with growth in expansions, where it is also more persistent. A nonlinear dynamic panel model estimates the threshold separating the two regimes at a quarterly GDP growth rate near minus one percent, which matches the recession threshold implied by Friedman’s plucking model of business cycles. Out-of-sample panel probit forecasts that use model uncertainty as the only predictor track recessions reasonably well at horizons up to one year, with the best performance at short horizons. A systematic increase in model uncertainty in recessions complicates the decision-making in real time, especially around the business cycle turning points.

**Key words:** model uncertainty, economic policy, output gap estimates

**JEL Codes:** E32, E37, E52, F44, C22

---

\*An earlier version of this paper was presented at the 99th annual Western Economic Association Conference, at The 10th Annual Conference of the Society for Economic Measurement (SEM2025), and at The 33rd Symposium on Nonlinear Dynamics and Econometrics (SNDE2026). We are grateful for the helpful comments provided by Luigi Donayre and Irina B. Panovska. All potential mistakes and imperfections remain our exclusive responsibility. This paper was funded by the European Union NextGenerationEU under Call IIP 581 2025. Data used in the paper is available on GitHub: <https://github.com/varcabic/gap-uncertainty>

<sup>†</sup>E-mail: jtica@efzg.hr, Office: +385.1.2383.162

<sup>‡</sup>E-mail: varcabic@efzg.hr, Office: +385.1.2383.229

<sup>§</sup>E-mail: vviljevac@efzg.hr, Office: +385.1.2383.229

# 1 Introduction

The output gap represents the relative deviation of real GDP from its potential level. Both the potential output level and the output gap are essentially unobservable. The literature has proposed several solutions, but there is no consensus on the best solution. However, output gap estimates are a fundamental tool for economic policy, influencing decisions in both monetary and fiscal policy, structural reforms, and economic forecasting and planning. Although output gaps are important, authors use various methods to estimate them, ranging from univariate [Hodrick and Prescott \(1981\)](#) (HP) and [Hamilton \(2018\)](#) filters to multivariate Structural Vector Autoregression (SVAR) and Bayesian VAR approaches based on Dynamic Stochastic General Equilibrium (DSGE) models ([Mazzi et al., 2017](#)).

Authors most often use one output gap technique (HP being the most popular), or when employing multiple output gap estimation methodologies, they often utilize various averaging techniques to derive the model-averaged output gap (MAOG). There is empirical evidence that MAOG has desirable properties and can even outperform individual output gaps used for its construction even when a finite number of output gaps are used for estimation ([Morley and Piger, 2012](#); [Morley, 2014](#); [Morley and Panovska, 2020](#)). However, there is a notable lack of academic attention to the dynamics and behavior of model uncertainty in the MAOG estimates (its standard deviation).

While there is a slight difference between the two approaches, we will stick to the term *model uncertainty* in the rest of the paper.

This paper (1) explores the properties of model uncertainty (disagreement) between various output gap estimation methodologies, (2) its connection to the GDP growth rate, and (3) its ability to forecast recessions. We measure model uncertainty as the cross-method standard deviation of output gap estimates, which captures the degree to which different estimation methods produce divergent readings of the cyclical position of the economy.<sup>1</sup> [Orphanides and van Norden \(2002\)](#) emphasize that the uncertainty attaches to the estimate of the output gap rather than to the model per se. Therefore, they would have called it estimate uncertainty,

---

<sup>1</sup>A coefficient of variation would also make sense for measuring uncertainty, where the standard deviation would be divided by the mean of the output gap, but it is difficult to calculate due to zero and near-zero values of the output gap which inflate the value of the coefficient of variation.

which is widely used in the real-time data literature. We use the term *model uncertainty* in the rest of the paper, but our measure is more accurately cross-method estimator disagreement. We adopt the shorter term for consistency with (Onatski and Williams, 2003). To calculate the model-averaged output gap (MAOG) and its standard deviation, ten empirical and three production function-based output gap methodologies are used.

Our goal is to investigate how model uncertainty changes over time and whether there are systematic changes related to GDP growth and recessions. The focus of our paper is to analyze this systematic relationship. We analyze how model uncertainty correlates with the GDP growth rate and how useful it is in forecasting recessions. Understanding the behavior of model uncertainty is as important as measuring output gaps, since the level of uncertainty is crucial for the decision-making process (Onatski and Williams, 2003). We analyze the dynamics of model uncertainty and use linear and nonlinear (threshold) dynamic panel fixed-effects models to estimate the effect of the GDP growth rate on model uncertainty. Furthermore, we use a battery of panel and time-series probit models to analyze the performance of model uncertainty in forecasting recessions.

The analysis is conducted in quasi-real-time using an expanding window, where one observation is added at a time so that each model is estimated using only the information that would have been available at time  $t$ . This mimics the environment in which policymakers and professional forecasters make decisions. The quasi-real-time approach differs from true real-time analysis only in that it uses final revised GDP data rather than the data vintage actually available at time  $t$  (Orphanides and van Norden, 2002). As documented in the literature, this distinction has limited practical consequences: quasi-real-time and true real-time output gap estimates are highly correlated, since the end-of-sample problem (addressed by the expanding window) is the dominant source of real-time unreliability, while the contribution of data revisions is comparatively minor (Orphanides and van Norden, 2002; Quast and Wolters, 2022). The quasi-real-time approach is therefore a natural choice when, as in our case, real-time GDP vintage archives are unavailable for all countries in the sample.<sup>2</sup>

In constructing the MAOG, we employ a total of thirteen output gap measures. For that

---

<sup>2</sup>Supplementary results based on full-sample output gap estimates are provided in Appendix C for comparison.

purpose, we use data on GDP, the employment rate, the unemployment rate, and the GDP deflator to estimate the output gap using ten different univariate and multivariate methods. In addition, we incorporate production function-based output gap estimates published by the European Commission, IMF, and OECD. We use quarterly data spanning the period from 2002Q1 to 2024Q4 for 25 EU countries and the United Kingdom.<sup>3</sup>

To analyze the relationship with GDP growth, we use linear and nonlinear panel data models. In addition to estimating a linear dynamic panel fixed-effects model, we employ a threshold approach to partition the sample into two regimes. We use the GDP growth rate as a threshold variable to examine how the effect of the business cycle on uncertainty varies across different phases (e.g., recessions versus booms). In addition, MAOG and the absolute value of GDP growth are used as alternative threshold variables.

For recession forecasting, we use panel and time-series probit models with a single independent variable – model uncertainty measured by the standard deviation of the output gap. We deliberately employ very simple models because the objective is not to conduct a forecasting horse race or identify the best-performing model. Instead, the goal is to demonstrate that model uncertainty possesses properties that are useful for forecasting recessions. We perform out-of-sample forecasting exercises for different models and compare their AUROC values. We also compare the forecasting performance of the model based on the standard deviation with that of a model that uses the average output gap as the sole predictor.

The paper contributes to the literature by analyzing the dynamics of output gap model uncertainty over the business cycle. Our findings indicate that uncertainty is not stable over time. It tends to be low and relatively constant during expansions but rises systematically during recessions. This cyclical pattern has direct implications for the precision of output gap estimates, particularly during downturns, when an accurate assessment of the current state of the economy is most consequential for policy.

In the linear dynamic panel fixed-effects model, we observe a strong negative effect of the GDP growth rate on uncertainty, which exhibits high persistence. Using a nonlinear threshold model, we find that the impact of growth rates on uncertainty is robust and significantly

---

<sup>3</sup>The sample includes all current EU member states except Ireland and Malta due to data availability and measurement issues, particularly concerning Ireland.

negative during deep slumps, while remaining indistinguishable from zero throughout the remainder of the sample. Our results lend support to the proposition that a nonlinear model structure provides a better explanation of the data than linear alternatives for many macroeconomic variables (Bradley and Jansen, 2017; Morley and Panovska, 2020; Huang et al., 2021).

Output gap model uncertainty is also a meaningful predictor of recessions. Estimated recession probabilities are intuitive, track actual recession episodes well, and could serve as a practical early warning indicator. In the panel setting, the standard deviation-based model does not systematically outperform the model that uses the average output gap, but in the dynamic fixed-effects specification the two are statistically indistinguishable, and in certain countries the standard deviation model offers a clear advantage. Taken together, the results suggest that output gap model uncertainty varies systematically over the business cycle and contains valuable information for recession forecasting.

An important practical feature of the proposed forecasting models is that all right-hand side variables are available in real time. The standard deviation of output gap estimates can be computed as soon as a new vintage of GDP data becomes available and the estimates are updated across methods, with no requirement for data subject to substantial revision or publication lags. The model is therefore operationally viable as a short-horizon early warning tool for national authorities or international institutions that monitor recession risk in real time.

The remainder of the paper is organized as follows. Section 2 summarizes the existing literature, while Section 3 provides an overview of the methodology and discusses the data. In Section 4 we present the empirical results, and Section 5 concludes.

## 2 Survey of Literature

The output gap is an unobservable variable. At the same time, it is a highly important theoretical, applied, and economic policy concept<sup>4</sup>. The literature has proposed many different

---

<sup>4</sup>For example, Cattelan and Nandwa (2024) show that output gap uncertainty can affect fiscal policy responses to different shocks. When policymakers are concerned about output gap uncertainty, they tend to overestimate shocks and respond more strongly, at the cost of public debt spiking in the short run and

ways of measuring the output gap (Mazzi et al., 2017), yet there is no consensus regarding the best or most appropriate measure (Morley and Panovska, 2020; Arčabić et al., 2024; Kamber et al., 2018).

The output gap can be estimated using a wide variety of different methods, which often produce different estimates. The simplest methods are univariate filters that use only real GDP, including the Hodrick and Prescott (1981) (HP) filter, Baxter and King (1999) filter, Christiano and Fitzgerald (1999) filter, Beveridge and Nelson (1981) decomposition, unobserved components (UC) models (Harvey, 1985; Clark, 1987), and the more recent Hamilton (2018) filter and its modification by Quast and Wolters (2022), each with its own strengths and weaknesses (see Mazzi et al., 2017; EU Independent Fiscal Institutions, 2020). Beyond these, multivariate approaches exploit additional macroeconomic information, for example by augmenting the HP filter with Phillips curve, Okun’s law, or capacity utilization relationships, as in Laxton and Tetlow (1992), using multivariate UC models and multivariate Beveridge–Nelson trends (Sinclair, 2009), or estimating gaps from structural VAR models along the lines of Blanchard and Quah (1989) and Bayoumi and Eichengreen (1992). Finally, institutional and model-based frameworks often rely on Cobb–Douglas production functions, as in the European Commission and OECD methodologies (Havik et al., 2014; Chalaux and Guillemette, 2019), or on dynamic stochastic general equilibrium (DSGE) models, which provide yet another class of multivariate output gap measures (see Segal, 2017).

Finally, empirical researchers have also struggled to agree on the very nature of the output gap, for example whether it is symmetric (linear) or asymmetric (nonlinear). Symmetric, or linear, output gaps typically fluctuate around potential output, being equally often above and below the potential level. Asymmetric, or nonlinear, output gaps, by contrast, tend to remain close to potential output, with occasional sharp declines associated with recessions (Hamilton, 1989; Kim and Nelson, 1999; Kim et al., 2005; Donayre and Panovska, 2021; Dupraz et al., 2025).

Given the absence of a clearly superior single method, several papers propose combining estimates across methods into a model-averaged output gap (MAOG). Morley and Piger (2012) introduce a Bayesian-weighted MAOG for the US and show it possesses the theor-

---

stabilizing in the long run.

retically desirable properties of a negative correlation with unemployment and a positive correlation with capacity utilization. [Morley \(2014\)](#) extends this approach to 13 economies using an unweighted average and identifies an additional desirable property: a negative correlation with future growth, consistent with the closing of the gap. [Morley and Panovska \(2020\)](#) confirm these results across 10 industrialized economies, while [Arčabić et al. \(2024\)](#) introduce a Phillips curve-weighted MAOG and show that business cycle synchronisation results for EU countries depend significantly on the method used, thereby challenging earlier findings of [Massmann and Mitchell \(2005\)](#). [Guérin et al. \(2015\)](#) find that model averaging improves the reliability of potential output estimates in real time, while [Hagelund et al. \(2018\)](#) show that the MAOG outperforms individual gap estimates in forecasting GDP growth, unemployment, and inflation.

These papers suggest that MAOG can be a superior measure of the output gap. However, it is important to point out that none of the previously mentioned papers study the uncertainty associated with the model-averaged gap or use it in further analysis, even though output gap estimates have been shown to vary substantially across methods and yield different conclusions (see [Arčabić and Olson, 2019](#); [Kumar, 2026](#)). Nevertheless, for every mean, a standard deviation can be calculated as a measure of dispersion. Accordingly, for the model-averaged output gap, the standard deviation across estimates can be interpreted as a measure of model uncertainty. This is the idea we explore in this paper. This standard deviation, as a measure of model uncertainty, may not be stable (equal) over time. Consequently, the representativeness of the MAOG may vary over time and over the course of the business cycle, potentially indicating that the MAOG is not an equally useful measure of the gap at all stages of the business cycle. A systematic pattern in which the standard deviation increases during recessions and declines during periods of economic growth may contain information useful for economic policy and recession forecasting ([Cattelan and Nandwa, 2024](#)). To our knowledge, no existing paper examines the properties of the standard deviation (model uncertainty) associated with the MAOG, even though this issue has been identified as important in the literature ([Kumar, 2026](#); [Cattelan and Nandwa, 2024](#)). This is the gap in the literature that the present paper seeks to fill.

## 3 Methodology and Data

### 3.1 Methodology

#### 3.1.1 Output Gap Estimation

To estimate output gaps, we employ a total of thirteen output gap estimation methodologies. All output gap measures are standard in the literature, and most of them are straightforward to implement in standard statistical or econometric software. We intentionally use such off-the-shelf approaches to address the model uncertainty associated with the most popular output gap measures.

The first ten gaps are estimated from quarterly real GDP data. We begin with the [Hodrick and Prescott \(1981\)](#) (HP) filter, the most commonly used univariate filter in the literature, and two band-pass filters: [Baxter and King \(1999\)](#) and [Christiano and Fitzgerald \(1999\)](#). We also employ a modified Beveridge–Nelson decomposition following [Kamber et al. \(2018, 2025\)](#), the original [Hamilton \(2018\)](#) filter and its modification by [Quast and Wolters \(2022\)](#), and a UC-0 unobserved components model following [Harvey \(1985\)](#) and [Clark \(1987\)](#). Finally, we estimate three structural VAR (SVAR) models with long-run restrictions in the spirit of [Blanchard and Quah \(1989\)](#), each including quarterly real GDP growth alongside one of the following variables: the employment rate, the unemployment rate, or the inflation rate. In each SVAR, potential output is recovered as the level of output that would prevail in the absence of demand shocks using historical decomposition.<sup>5</sup>

The remaining three gaps are taken from the official estimates of the European Commission, the OECD, and the IMF. The European Commission and the OECD both rely on a Cobb–Douglas production function approach, estimating potential output via the natural rate of unemployment at which unit labor costs (European Commission) or core inflation (OECD) are stable; full methodological details are provided in [Havik et al. \(2014\)](#) and [Chalaux and Guillemette \(2019\)](#), respectively. The IMF does not apply a single unified methodology across countries; instead, analysts select the approach considered most appropriate for each country, drawing on univariate and multivariate filters, the production function, and expert judgment ([De Masi, 1997](#); [Barkema et al., 2020](#)).

---

<sup>5</sup>For further details on each estimation method, see Appendix [A](#).

Since the institutional estimates are published at the annual frequency while the remaining ten gaps are quarterly, we convert the annual institutional gaps to a quarterly frequency. Specifically, we first recover annual potential GDP implied by each institution’s gap estimate, then interpolate it to quarterly frequency using a quadratic-match sum procedure, which preserves the annual sum of potential output and ensures the interpolated series is consistent with the original institutional estimates. The quarterly output gap is then computed as the percentage difference between actual seasonally adjusted real GDP and the interpolated quarterly potential GDP. This yields thirteen output gap series used throughout the analysis.

### 3.1.2 Quasi-real-time and final output gaps

Following [Orphanides and Norden \(2000\)](#), we estimate and use both the quasi-real-time and final output gaps. The quasi-real-time output gap estimate for period  $t$  is based on time series which as the last data point includes the data for period  $t$  (the expanding window method). That is, one attempts to historically reconstruct the gap estimate that could have been obtained for period  $t$  right after all the data required to estimate the gap for period  $t$  have been published. The data for later periods (after period  $t$ ) are not used for this estimation because they were not available in real-time when it was first possible to estimate the gap for period  $t$ . The difference between real-time and quasi-real-time estimates is that real-time estimates use original (vintage) data, while quasi-real-time estimates use the most recent (currently available) vintages of data, which were possibly revised compared to the vintage data (for example, GDP figures get revised often, etc.). Both quasi-real-time and real-time output gap estimates are essentially obtained using the expanding window method. For example, if one is trying to estimate the (quasi) real-time output gap for France for 2012Q3, the last data used for the estimation is the data for 2012Q3.

On the other hand, final estimates of the output gap use all currently available data to estimate the gap for period  $t$ . For example, if one is trying to estimate the output gap for France for 2012Q3, all of the currently available data will be used to estimate that gap, for instance, the data up until 2023Q4.

### 3.1.3 Model uncertainty and model averaged output gap (MAOG)

In total we employ up to 13 different methodologies to estimate the  $K$  output gaps  $c_{itk}$  for a specific country.<sup>6</sup> Due to data availability and the credibility of the estimated gaps in some countries we use smaller number of gaps  $K < 13$ .

The total number of models employed can be much higher, [Morley and Panovska \(2020\)](#) for example use 13 linear and 14 nonlinear models in order to estimate MAOG. In this paper, we use 13 methodologies, but we can easily expand the model along that dimension.

Since we are focused on output gaps standard deviation, we use only symmetric, or linear output gaps. The literature also recognizes an important role of asymmetric or nonlinear output gaps ([Morley and Piger, 2012](#); [Morley and Panovska, 2020](#); [Arčabić et al., 2024](#)). However, combinations of both symmetric and asymmetric output gaps would increase the dispersion by construction, and thus affect the measured standard deviation. Therefore, in this paper we do not pursue this avenue.

After  $K$  gaps have been estimated, typically different averaging techniques are employed to calculate a single estimate of the model-averaged output gap  $c_{it}$ . Authors usually use the unweighted average ([Morley and Panovska, 2020](#)), weighted average based on the Bayesian approach ([Morley and Piger, 2012](#)), or a theoretical justification such as the Phillips curve relationship vis a vis the inflation rate ([Arčabić et al., 2024](#)).

In the paper we estimate the model average output gap (MAOG) as the equally weighted average of  $K$  output gaps for each country and time period:

$$c_{it} = \frac{1}{K} \sum_{k=1}^K c_{itk} \quad (1)$$

where  $k = 1...K$  number of employed output gap methodologies, and  $c_{itk}$  is the estimated MAOG in time  $t$  and country  $i$ .

Although the calculation and robustness of MAOG approach has been extensively researched in the literature, as far as authors are aware the level of model uncertainty is rarely employed in the analysis of output gap. Formally, we define uncertainty as the standard

---

<sup>6</sup>HP filter, Hamilton filter, Quast and Wolters filter, Baxter-King filter, Christiano-Fitzgerald filter, Kamber, Morley and Wong's refined BN filter, UC-0 method, three SVAR models and the output gap estimates from European Commission, OECD and IMF.

deviation of estimated gaps  $c_{itk}$  vis-à-vis the MAOG  $c_{it}$ .

$$\delta_{it} = \sqrt{\frac{\sum_1^K (c_{itk} - c_{it})^2}{K}} \quad (2)$$

Every estimate of the MAOG  $c_{it}$  has a different level of uncertainty (standard deviation) in each country and time period, denoted as  $\delta_{it}$ .

### 3.1.4 Linear and threshold dynamic panel model

Many authors are well aware that model uncertainty exists and that it can be quite volatile through time and space, but the key question in this paper is how to utilize the information about the level of uncertainty in the process of gap estimation and eventually in policy making. In this paper, we analyze the connection between GDP growth rate, MAOG, and model uncertainty. We employ linear and nonlinear dynamic panel model ([Chan, 1993](#)) in order to investigate the issue.

We start with one way and two way linear dynamic fixed effect panel model:

$$\delta_{it} = \alpha_i + \gamma_t + \rho\delta_{it-\tau} + \beta\Delta y_{i,t} + e_{1i,t} \quad (3)$$

and proceed with the following form of the threshold two way dynamic panel data model

$$\delta_{it} = \alpha_i + \gamma_t + \begin{cases} \rho_1\delta_{it-\tau} + \beta_1\Delta y_{i,t} + e_{1i,t} & \text{if } q_{i,t-d} \leq \theta \\ \rho_2\delta_{it-\tau} + \beta_2\Delta y_{i,t} + e_{2i,t} & \text{if } q_{i,t-d} > \theta, \end{cases} \quad (4)$$

where  $\delta_{it}$  is the standard deviation between various gap estimates (a measure of uncertainty),  $\tau$  is the lag of dependent variable, we estimate model without lag dependent variable, with  $\tau = 4$  and with two lag dependent variables  $\tau = 4$  and  $\tau = 1$ ;  $\Delta y_{i,t}$  denotes the GDP growth rate where y stands for  $\log(\text{GDP})$ ,  $q_{i,t}$  is the threshold variable. We estimate model with GDP growth rate, MAOG and absolute value of GDP growth rate as threshold variables.  $\theta$  is the estimated threshold value/parameter;  $\alpha_i$  and  $\gamma_t$  capture cross-sectional and time fixed effects, respectively; and  $d$  is the delay parameter. We use both  $d = 0$  and  $d = 1$  to address

the problem of intertemporal links between regimes.

In order to estimate threshold value  $\theta$  we grid search for the best sample split by minimizing loss function represented by root mean square error parameter. The grid search for the threshold value is performed through all the observations of threshold variable between 10th and 90th percentile.<sup>7</sup>

### 3.1.5 Probit models and recession forecasting

We also use the output gap standard deviation ( $\delta_{i,t}$ ) to forecast recessions. Recessions are identified using a standard rule of thumb, where a recession is called if the Y-o-Y growth rate of real GDP is negative for two consecutive quarters. This binary indicator (1 for recession quarters, and 0 otherwise) serves as the dependent variable in all models estimated below.

We use panel probit models on a balanced panel of 25 EU member states plus the United Kingdom observed at quarterly frequency. The full estimation sample begins in 2002:Q1; the out-of-sample forecast evaluation window covers 2012:Q1–2024:Q4, where the cutoff is chosen to leave a sufficient estimation window that includes the Global Financial Crisis while preserving more than a decade of genuine out-of-sample observations spanning the euro-area sovereign debt crisis, the COVID-19 recession, and the 2022–2023 energy-price shock.

The baseline specification is a panel probit with an unobserved country effect  $\alpha_i$

$$P(R_{i,t+h} = 1 \mid \mathbf{x}_{i,t}, \alpha_i) = \Phi(\mathbf{x}_{i,t}'\boldsymbol{\beta} + \alpha_i), \quad (5)$$

where  $\Phi(\cdot)$  is the standard normal cumulative distribution function,  $\boldsymbol{\beta}$  is the coefficient vector, and the forecast horizon is  $h \in \{1, 2, 3, 4\}$  quarters. Forecasts are produced directly rather than by iteration, so a separate model is estimated for each  $h$ . The regressor vector  $\mathbf{x}_{i,t}$  contains an intercept and the output gap model uncertainty measure  $\delta_{i,t}$  as its only non-constant element. In the static specification  $\mathbf{x}_{i,t} = (1, \delta_{i,t})'$ ; in the dynamic specification the lagged recession indicator  $R_{i,t}$  is added to capture state persistence.

We estimate equation (5) under two treatments of  $\alpha_i$ . In the random effects (RE) probit,  $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$  is assumed independent of  $\mathbf{x}_{i,t}$  and the likelihood is evaluated by adaptive

---

<sup>7</sup>For the robustness check we also perform grid search between 5th and 95th percentile (see Figures 5 and B.1 for details).

Gauss–Hermite quadrature. In the fixed effects (FE) specification, we estimate a pooled probit with a full set of country indicator variables. We refer to this as an "FE probit" for brevity, but note that with finite  $T$  the unit-dummy MLE suffers from the incidental parameters problem (Neyman and Scott, 1948): the slope estimator is asymptotically biased of order  $O(1/T)$  and the bias does not vanish under fixed- $T$  asymptotics (Greene, 2004). With  $T \approx 92$  quarterly observations per country the bias is expected to be small. Inference throughout uses standard errors clustered at the country level, which is robust to arbitrary within-country serial correlation and to the conditional heteroskedasticity induced by the probit link. Combining the two static and two dynamic specifications yields four models: RE-static, RE-dynamic, FE-static, and FE-dynamic.

In the RE-dynamic specification the lagged outcome  $R_{i,t}$  is by construction correlated with  $\alpha_i$ , giving rise to the initial conditions problem (Heckman, 1981). We address this using the auxiliary-model approach of Wooldridge (2005), which parameterizes the conditional distribution of the unobserved effect as

$$\alpha_i \mid R_{i,1}, \bar{\delta}_i \sim \mathcal{N}(a_0 + a_1 R_{i,1} + a_2 \bar{\delta}_i, \sigma_a^2), \quad (6)$$

where  $R_{i,1}$  is the initial recession state and  $\bar{\delta}_i$  is the within-country time-series mean of the uncertainty regressor. Substituting into (5) and integrating out  $\alpha_i$  yields a standard RE probit in which  $R_{i,1}$  and  $\bar{\delta}_i$  enter as additional regressors, so that the unobserved effect is modeled conditional on the initial state rather than assumed independent of the dynamic history. The static RE specification contains no lagged dependent variable and is therefore free of the initial conditions problem.

Our probit models are deliberately simple, because the idea is not performing a forecasting horse race and finding the best model, but showing that model uncertainty has properties that are useful in forecasting recessions. The choice of probit model is not by chance. Chalaux et al. (2025) show that an algorithm that selects appropriate probit models, matches the performance of Random Forests in out-of-sample predictions, including the period of the Global Financial Crisis. Such performance is attributed to a *wisdom of crowds* property, as predictions from probit equations are averaged.

Out-of-sample forecasting is based on a recursive expanding window scheme. For each

quarter  $t$  from 2012:Q1 to 2024:Q4, the model is re-estimated using all panel observations up to but excluding quarter  $t$ , and the resulting coefficient estimates are used to compute the predicted recession probability for quarter  $t$ . The initial training sample thus spans from 2002:Q1 to 2011:Q4, and it expands by one quarter at a time. This recursive procedure replicates the information constraints of a real-time forecasting exercise, since each forecast uses only data that would have been available at the time it was produced.

The second part of the analysis estimates a probit model for each of the 26 countries, in a time-series setting. As in the panel case, static and dynamic probit specifications are estimated with lags from one to four quarters. In the time series setting, there are no fixed effects, and therefore, there is no potential bias as in a panel setting.

Model forecasting performance is evaluated using the area under the receiver operating characteristic curve, AUROC. The ROC curve plots the true positive rate against the false positive rate as the classification threshold for the predicted recession probability is varied continuously from zero to one, and the AUROC is the integral of this curve. An AUROC of 0.5 indicates no discriminatory power relative to a random walk model, values around 0.7 are generally considered acceptable, values around 0.8 are considered good, and values above 0.9 indicate excellent discrimination. To assess whether differences in AUROC values across different specifications are statistically significant, we use the DeLong test ([DeLong et al., 1988](#)). The DeLong test compares two correlated ROC curves estimated on the same observations and tests the null hypothesis that their AUROC values are equal.

## 3.2 Data

Since every output gap estimation method requires data for the real GDP of the country whose gap is being estimated, data on seasonally and calendar-adjusted real GDP (in euros) was taken from two databases, the Eurostat database ([Eurostat, 2025c](#)) and the OECD database ([OECD, 2025b](#)). The frequency of the data is quarterly, and the reason why these two databases are combined is to obtain the longest possible series for all countries in the sample. The real quarterly GDP series were expressed in logarithmic form before applying each univariate method. The difference between the logarithms of actual and potential GDP represents the output gap. Real GDP data for the 26 countries in our sample is shown in

Figure A.1, and Figure A.2 displays the real GDP growth rates.

A similar approach was used when retrieving the data on the unemployment rate and the employment rate that are needed to estimate the output gap using the Blanchard-Quah SVAR model. To obtain the longest possible series of data on the unemployment rate, for some countries, we use quarterly data for the age group 15+ from the OECD database (OECD, 2025c), while for the other countries we use the data for the age group 15-74 obtained from the Eurostat database (Eurostat, 2025e) and the Eurostat database with historical data (Eurostat, 2025f). The use of different age groups does not affect our results because our research is not based on mutual comparison of gaps between countries. Quarterly data on the employment rate in the age group 15-64 was taken from Eurostat, combining current (Eurostat, 2025a) and historical (Eurostat, 2025b) data depending on the availability of data for each country in the sample. The data for the UK's employment rate was obtained from the OECD's database (OECD, 2025c). The quarterly GDP deflator (price) data needed to estimate the output gap using the Bayoumi and Eichengreen SVAR model was taken from Eurostat (Eurostat, 2025d) and the UK's Office for National Statistics (Office for National Statistics, 2025). The employment and unemployment rate data is shown in figure A.3, while figure A.2 displays the GDP deflator rate of inflation.

When it comes to real-time gaps estimated by the European Commission, the data was taken from the autumn editions of the publication "Economic Forecast" published in the period from 2003 to 2025, from which we obtained the values of the gaps for the period from 2002 to 2024 (European Commission (2025)). We take the first publicly available output gap estimate as the real-time gap estimated by the European Commission. The first year for which the European Commission published gap estimates was 2002 for a wider set of European Union countries, hence this year was taken as the starting year. Final gaps estimated by the European Commission were obtained from the same publication and from the last vintage which published output gap estimates for a specific year. A similar approach was used to obtain the real-time and final output gaps estimated by the IMF and the OECD. The IMF publishes gap estimates as a part of the "World Economic Outlook" publication, and these gap estimates are available in the IMF database (IMF, 2025). Estimates of output gaps published by the OECD were taken from the OECD database and the publication

”Economic Outlook” ([OECD, 2025a](#)).

## 4 Results

The results section is divided into three parts. First, we present the results of output gaps estimation and model uncertainty. Secondly, we provide estimates of linear and nonlinear dynamic panel models. Third, we perform different forecasting exercises, showing that the model uncertainty is a meaningful signal in recession forecasting, and that it includes a valuable information.

### 4.1 Output gaps and model uncertainty

Figure 1 presents all estimated gaps used in the analysis for each country. All depicted gaps are estimated for the quasi-real-time sample.<sup>8</sup> As shown in Figure 1, output gaps are similar, but there are periods when the divergence between various methodologies is large and persistent. In a few countries, such as the UK, output gap measures are highly synchronized. But in the most countries, there are substantial disagreements between output gap measures, even about the current state of the economy.

To see if such disagreements are random or systematic, we calculate the model uncertainty measured by the standard deviation. Results for the quasi-real-time sample are presented in Figure 2. The Figure shows three versions of standard deviation based on the quasi-real time. The first includes all thirteen output gaps, the second excludes institutional output gaps, while the third excludes [Hamilton \(2018\)](#) filter and UC-0 output gap.<sup>9</sup> Periods of negative Y-o-Y GDP growth are shaded.

In many countries in our sample, shaded periods are usually periods of higher model uncertainty. Three major crisis are common to almost all the countries: the Global Financial Crisis (GFC) of 2008-2010, the Sovereign debt crisis in the Eurozone (2010-2012) and the Covid recession of 2020-2021. A substantial increase in the model uncertainty is present

---

<sup>8</sup>Due to the differences in data availability, MAOG was calculated using a smaller number of gaps (fewer than 13) for some countries. Full sample output gap results are presented in Figure C.1.

<sup>9</sup>Figure C.2 presents the comparison between model uncertainty estimated for quasi-real-time gaps and final gaps.

during the GFC and Covid recession. In most countries, the model uncertainty was moderate during the Sovereign debt crisis, probably due to the very nature of the crisis. Namely, it coincident with the end of the GFC, and for many countries it had a form of a double-dip recession or it was a single long lasting recession, such as in Greece and Croatia.

The general picture that arises is that there is a systematic increase in uncertainty in recession periods. Such a systematic increase in uncertainty remains regardless of the exclusion of different output gap measures. This can be clearly seen in Figure 2 as the differences between all three versions of standard deviation are small. Therefore, our measure of model uncertainty is not mainly driven by one or a few output gap estimates, but rather it is a feature of the data. In other words, such systematic increase in uncertainty in recession periods is not by chance, but by design.

To further highlight the point that our model uncertainty measure is not driven by a few output gap measures, Figure 3 shows the variance decomposition of our model uncertainty measure, which includes all 13 measures of the output gap. The figure is shown in a stacked bar chart which sums to 100%. The interpretation is straightforward: a percentage of variation in model uncertainty explained by each particular output gap measure.

More specifically, we decompose the cross-method variance of the output gap into method-specific components. For each country and each quarter, the total variance is computed as the mean squared deviation of the individual output gap estimates from their cross-method average. The contribution of method  $k$  to the total variance in quarter  $t$  is defined as:

$$\psi_{kt} = \frac{1}{K} (c_{kt} - \bar{c}_t)^2 \quad (7)$$

where  $c_{kt}$  is the output gap estimate from method  $k = 1, \dots, K$ , and  $\bar{c}_t = \frac{1}{K} \sum_{k=1}^K c_{kt}$  is the cross-method mean. By construction,  $\sum_{k=1}^K \psi_{kt} = \sigma_t^2$  in every quarter, so the shares  $\omega_{kt} = (\psi_{kt}/\sigma_t^2) \times 100$  sum exactly to 100. We average country-level shares over time and present means.

The decomposition helps determine which output gaps dominate in our model uncertainty measure. In quasi-real time, particular output gap measures do not take disproportionately large shares, and the variance decomposition suggests our model uncertainty measure is not driven by a single or a few output gap measures. Rather, the shares are widely distributed.

Results vary between the countries, but [Hamilton \(2018\)](#) filter, [Quast and Wolters \(2022\)](#) filter, UC-0 and all three specifications of SVAR models have the largest shares. However, institutional output gap measures calculated by the European Commission, OECD, and IMF are also important. Other output gap measures also have non-trivial shares. Only [Hodrick and Prescott \(1981\)](#) filter seems to have very small, almost a negligible share in the variance decomposition, which is consistent with its mechanical smoothness properties and reinforces the critique in [Hamilton \(2018\)](#).

## 4.2 Nonlinear relationship to GDP growth

To formally test for the connection between the economic slowdown and the rise in model uncertainty, we estimate several econometric models. Tables 1 to 4 present the estimated results of linear and nonlinear (threshold) dynamic panel models (equations 3 and 4). All the results are for the quasi-real-time output gaps.<sup>10</sup>

Table 1 presents the results for the linear version of equation 3. We start with the linear model to highlight the degree of differences in estimated parameters and p-values between the linear model and the threshold models with two regimes.

The first three models in Table 1 show the results for one-way fixed effects model, while the last three estimated models show results for the two-way fixed effects model.

The effect of the GDP growth rate  $\Delta y$  on model uncertainty  $\delta$  is negative and statistically significant at 1% in all six models. Both lags of the dependent variable are statistically significant and have a positive effect in all models, but the model with both lags of the dependent variable in the two way fixed effect panel has the best fit overall (The last three rows in Table 1 represent root mean square error (RMSE), Aikike’s (AIC) and Schwarz’s Bayesian information criteria (BIC)).

This implies that, on average, positive economic conditions are correlated with smaller model uncertainty between various output gap estimation methodologies. Furthermore, a robust estimate of the effect of the lagged dependent variable with coefficients between .51 and .80 implies that model uncertainty is highly persistent.<sup>11</sup>

---

<sup>10</sup>Estimates for the output gaps using full sample are presented in Table C.1.

<sup>11</sup>In the appendix we provide analogous results for the final output gap estimates. The results are slightly

Next, we use nonlinear panel data models to analyze the asymmetric relationship between GDP growth and model uncertainty. Table 2 and Figure 4 show the results of the dynamic threshold panel two-way fixed effects model with GDP growth rate as the threshold variable  $q_{it-d}$ . This is our benchmark model, and we compare all other model to this one. For nonlinear panel data models, we skip one-way fixed effects models for brevity.

Results for six estimated models are presented. The first model is estimated without the lagged dependent variable, the second model with one ( $t - 4$ ), and the third model with two lags of the dependent variable ( $t - 1$  and  $t - 4$ ). Models 1 through 3 are estimated with delay parameter  $d = 0$ , while in models 4 through 6 we repeat first three estimates, but with delay parameter set to  $d = 1$ .

In Table 2, the suffix \_1 denotes the estimated coefficients for regime 1 that represents the sub-sample of data below the threshold value  $\theta$ , while the suffix \_2 represents coefficient values for the sub-sample above the threshold value  $\theta$ .

The results are most intriguing. We find a strong evidence of asymmetric relationship between GDP growth and model uncertainty. In the lower regime below the threshold, there is a strong negative relationship between GDP growth and model uncertainty, and no relationship at all in the upper regime above the threshold. In all six estimated models, over all combinations of dependent variable lags and delay parameters, the effect of GDP growth rate on model uncertainty in regime 1 ( $< \theta$ ) is negative and statistically significant, with values between  $-.3$  and  $-.13$ .

On the other hand, in regime 2 (above the threshold value), the effect of GDP growth rate on model uncertainty is not statistically different from zero in any of the six models. If the GDP growth rate is above the endogenously estimated threshold value, there is no evidence of a correlation between growth and model uncertainty.

We also provide formal test results for the p-value of the Wald test in row *Wald\_pval* at the bottom of Table 2. Null hypothesis of the Wald test that coefficients for GDP growth rates in both regimes are equal  $dy_1 = dy_2$  is firmly rejected in all models.

This suggests that an increase in model uncertainty is associated with negative GDP

---

different but do not change our major conclusions. Table C.1 shows the results of pooled OLS, one-way fixed effects, and two-way fixed effects models for final sample output gaps.

growth, which has two implications. First, during recessions, policy makers cannot be sure of the depth of the recession, due to increased uncertainty. It makes decision making in real time more difficult, and wrong or harmful policy decisions are more likely. Second, an increase in model uncertainty can be an useful signal in forecasting recessions, which we analyze further in section 4.3.

When it comes to the persistence of the model uncertainty variable, the behavior of lags in the dependent variable is quite similar in both regimes, but the estimated coefficients are much larger in regime 2 (above the threshold value). The higher parameter persistence in the upper regime indicates that model uncertainty decreases more slowly during expansions and mild slowdowns than during severe contractions. This asymmetry suggests that sharp recessionary contractions generate a strong common signal that disciplines estimates across methods, whereas the absence of such a signal during normal times allows cross-method disagreement to linger for extended periods. The higher persistence of model uncertainty in the upper regime is particularly relevant for economic policy. It implies that elevated disagreement across output gap measures is not quickly self-correcting during expansions. This complicates the decision making in real time, especially around the business cycle turning points.

When it comes to estimated threshold values that maximize the root mean squared error (RMSE) by splitting the data sample, results are presented in row "Tr\_value" of Table 2. Estimated thresholds are between 0.02 in model 1 and  $-1.01$  in models 4 to 6, but the root mean squared error is minimized in model 3 with two lags of the dependent variable with the threshold value of  $-0.99$ . We also include AIC and SBC indicators at the bottom of the table in order to control for model parsimony, but even after penalties for additional regressors, the model 3 still have best fit. Threshold values in models 3 to 6 are very close to each other.

Figure 4 presents the results of the grid search for the threshold value for each of the six models from Table 2. In each sub-figure we present the estimates of the mean squared errors (vertical axis) for each possible threshold value between the 11th and 89th percentile. The threshold value that minimizes RMSE in each of the six models (from Table 2) is marked with a vertical line.

Overall, model 3 has the best fit (Figure 4e) and the endogenously estimated threshold

value that minimizes the loss function is GDP growth rate of  $-0.993\%$ . According to our model, a drop in GDP higher than  $1\%$  creates an environment in which the GDP growth rate is negatively correlated with model uncertainty.

The estimated threshold of  $-0.99\%$  GDP growth can be related to asymmetric nature of business cycles described by the [Friedman \(1993\)](#) "plucking" model. A contraction of this size is consistent with the onset of a proper recession, and not just a mild slowdown or technical recession. In our sample of EU countries, observations falling in the lower regime correspond almost exclusively to well-documented crisis episodes: the Global Financial Crisis of 2008–2009, the euro area sovereign debt crisis, and the COVID-19 recession. Mild recessions and growth slowdowns, by contrast, remain within the upper regime. This pattern connects naturally to the plucking model of business cycles proposed by [Friedman \(1993\)](#), empirically modeled by [Kim et al. \(2005\)](#) and formalized in recent work by [Dupraz et al. \(2025\)](#) and [Donayre and Panovska \(2021\)](#), in which the economy operates near its potential most of the time and is occasionally "plucked" downward by larger negative shocks. In this sense, the estimated threshold can be interpreted as identifying the boundary between the two regimes of the plucking model: the upper regime captures normal business cycle variation around potential output, while the lower regime captures the episodic, asymmetric contractions that define the "plucks" themselves.

The most threshold values in our model in [Table 2](#) are similar in magnitude. However, a caveat is warranted. Threshold parameters are often estimated imprecisely, and their statistical significance is difficult to test due to the wide confidence intervals typically associated with them and the well-known identification problems under the null of no threshold effect (see [Hansen, 2000](#); [Enders et al., 2007](#)). We therefore interpret the estimated threshold as indicative rather than definitive, and we emphasize that our conclusions rest on the qualitative difference in regime dynamics and not on the precise location of the threshold.

To visualize our main results, we present the scatter plot of model uncertainty vs. GDP growth rate for the entire panel in [Figure 5](#) and for each country in the panel in [Figure 6](#). The vertical line represents the estimated threshold value in model 3 that minimizes the loss function (Model 3 in [Table 2](#)).

Even without controlling for the persistence of dependent variables, observations in the

sub-sample below estimated  $\theta$  are in gray and the negative correlation is obvious. On the other hand, in the rest of the sample the correlation is not visible (Figure 5).

A similar result is visible for the majority of countries in Figure 6. Exceptions are Bulgaria (BG), Cyprus (CY) and Hungary (HU) above the threshold value where the correlation is positive, and Romania (RO) with weak correlation below the threshold value.

The negative slope that is very clearly pronounced below the 10th percentile of GDP growth rates (Figure 5) served as a motivation to expand the grid search to include the range between the 5th and the 95th percentile. In other words, a grid search between the 10th and 90th percentile may not do as good job of detecting the break as the grid search between the 5th and 95th percentile. This is because a number of recession periods in the overall sample is relatively low, and recessions are mainly concentrated below the 10th percentile.

The results for the models where the grid search is performed between the 5th and 95th percentile using the GDP growth rate as the threshold variable (Table 3) are similar to the results obtained when a grid search between the 10th and 90th percentile is used (Table 2). In other words, the results do not change if we allow more extreme threshold values. GDP growth still has a negative and statistically significant effect on uncertainty below the threshold, and the effect is not statistically significant above the threshold, except in model (6), while model 3 still has the best fit. The null hypothesis of the Wald test that the coefficients for GDP growth rates are equal in both regimes is once again rejected in all six models. Figure B.1 presents the results of the grid search for the threshold value for each of the six models from Table 3. Threshold values in this model differ more than in our benchmark model presented in Table 2, ranging from  $-3.6$  in Model 3 to  $+3.7$  in Model 4. The threshold value of  $-3.6$  in Model 3 is even more supportive of the plucking model of business cycles (Friedman, 1993), as a growth rate of  $-3.6$  suggests severe recessions. However, as we have already noted, threshold values are not precise, and we tend to interpret them as indicative rather than definite.

A natural way to distinguish how model uncertainty correlates with positive versus negative GDP growth is to fix the threshold exogenously at  $\Delta y_{i,t} = 0$  rather than estimate it from the data. Table 4 reports the results of this exercise. We replace the estimated threshold parameter  $\theta$  with an indicator function  $\mathbb{I}\{\Delta y_{i,t} < 0\}$  that splits the sample into recessionary and

non-recessionary observations, leaving the rest of the specification in equation 4 unchanged. The results closely mirror those of our benchmark model in Table 2: the coefficient on GDP growth is negative and statistically significant at the 1% level across all six specifications in the regime below zero, while it is statistically indistinguishable from zero and economically negligible in the regime above. The Wald test rejects equality of the two regime coefficients in every specification. The qualitative finding that the negative association between GDP growth and model uncertainty is concentrated in contractions is therefore not an artefact of the data-driven threshold; it survives when the regime boundary is imposed at the natural economic cutoff of zero growth. The zero-threshold specification pools mild contractions with severe ones, yet the regime asymmetry persists, indicating that the negative correlation is not reserved to the deepest plucking episodes only.

In Appendix B we perform several robustness checks, presented in Tables B.1 to B.4. All robustness checks are based on nonlinear models and are organized as in the benchmark model.

Table B.1 presents the results of models estimated with the MAOG as a threshold variable  $q_{it-d}$  in equation 4. The results are robust: we find a negative and statistically significant coefficient next to GDP growth rate in the lower regime, and insignificant and close to zero coefficients in the upper regime. Only in model (5) the coefficient in the upper regime is negative and significant at 5%, but Wald test rejects equality of coefficients in all six specifications. Threshold values are again more dispersed than in the benchmark model, ranging from  $-2.45$  to  $1.63$ , with the threshold value of  $-2.15$  in the best fitting Model 3 according to RMSE and information criteria<sup>12</sup>.

To examine whether the asymmetry documented in our benchmark specification reflects the *magnitude* of GDP fluctuations rather than their *direction*, we re-estimate equation 4 using the absolute value of the GDP growth rate,  $|\Delta y_{i,t-d}|$ , as the threshold variable. If methodological disagreement across output gap measures is driven primarily by the size of the underlying GDP shock — irrespective of whether the economy is contracting or expanding — then the regime asymmetry should re-emerge under this specification, with both

---

<sup>12</sup>Visualization of the grid search procedure for each of the six models in Table B.1 is displayed in Figure B.2

regimes producing comparable coefficients on  $\Delta y_{i,t}$ . Conversely, if the asymmetry in the benchmark model genuinely reflects the sign of growth, the absolute-value threshold should fail to reproduce it. Table B.2 reports the estimated coefficients for the six model variants, and Figure B.3 displays the corresponding grid-search results.

The estimates support the second interpretation, e.g., its about the sign, not the magnitude. The coefficient on GDP growth is negative and statistically significant at the 1% level in both regimes across all six specifications, with magnitudes that are similar in absolute terms — for instance,  $-0.155$  versus  $-0.086$  in Model 1 and  $-0.058$  versus  $-0.020$  in Model 6. The Wald test fails to reject equality of the regime coefficients at the 5% level in five of the six specifications, with Model 6 the sole exception. The persistence of model uncertainty, captured by the lagged dependent variable, is likewise comparable across the two regimes. Taken together, these results indicate that splitting the sample by the magnitude of GDP changes does not generate the regime asymmetry observed in the benchmark specification: when contractions and expansions of similar size are pooled into the same regime, the negative association between growth and uncertainty is restored throughout. The asymmetric pattern documented in Table 2 therefore cannot be attributed to a mechanical artefact of large absolute changes in GDP. It reflects a genuinely sign-dependent relationship in which contractions, but not expansions of equivalent magnitude, increase disagreement across output gap estimation methods.

As a final robustness test, tables B.3 and B.4 include the results of the models where the model uncertainty is calculated without either institutional output gaps (Table B.3) or UC and Hamilton output gaps (Table B.4). The results are qualitatively the same and quantitatively very similar to the results in the benchmark model (Table 2). Once again, in all six estimated models the effect of the GDP growth rate on model uncertainty is negative and statistically significant only in lower regime. The Wald test rejects the null hypothesis of the equality of the estimated coefficients in the two regimes. Visualization of the grid search procedure for the models in these two tables is shown in Figures B.4 and B.5.

### 4.3 Forecasting recessions with model uncertainty

Our probit models show that output gap model uncertainty, measured as the standard deviation of real-time estimates, contains a valuable signal for recession forecasting in an out-of-sample setting, as we discuss below. The models are deliberately parsimonious, including only output gap model uncertainty as the single explanatory variable. The purpose of this exercise is not to identify the best-performing forecasting model in a horse race, but rather to demonstrate that model uncertainty possesses features that are useful for recession prediction. We show not only that it can beat random walk benchmarks, but that at shorter forecasting horizons the performance is, by conventional standards, quite good.

A natural limitation of this forecasting exercise is the length of the available sample. The full sample covers 2002:Q1 to 2024:Q4. We split it approximately in half, using 2002:Q1 to 2011:Q4 as the training sample and 2012:Q1 to 2024:Q4 as the out-of-sample evaluation period, which is implemented as an expanding window. The out-of-sample window is short by the standards of forecasting evaluation, and at most three recessions occurred within it for any given country. This limits our ability to draw strong conclusions about model performance on the basis of statistical significance alone, and it is a caveat that applies to all results discussed below.

Figure 7 presents out-of-sample AUROC values from the panel probit models. Results are reported for FE and RE specifications, in both static and dynamic form, at forecasting horizons  $h = 1$  to 4. The two horizontal reference lines mark AUROC values of 0.5, the threshold corresponding to a random classifier, and 0.7, a conventional threshold for a reasonably informative model. Across all specifications, dynamic models outperform their static counterparts, and FE models perform somewhat better than RE models.

At shorter horizons the models perform well. The dynamic FE and RE models at  $h = 1$  reach AUROC values of 0.87 and 0.86, respectively, and at  $h = 2$  these decline to 0.75 and 0.72. Both values comfortably exceed the 0.7 threshold and can be considered good. Static models perform worse, though the FE static model still reaches AUROC values of 0.63 at  $h = 1$  and 0.61 at  $h = 2$ , both above the random walk benchmark.

At longer horizons,  $h > 2$ , performance deteriorates, as expected, but most models continue to outperform the random walk. The exception is the RE static model, which falls

below 0.5 at  $h = 4$ . The FE models retain respectable performance even at longer horizons, with the dynamic FE model reaching AUROC values of 0.67 at  $h = 3$  and 0.60 at  $h = 4$ , and the FE static model reaching 0.58 at both  $h = 3$  and  $h = 4$ .

The monotonic decline in AUROC from  $h = 1$  to  $h = 4$  is expected, but the rate of decline and the level at which predictive content survives are informative. The dynamic FE model retains an AUROC of 0.60 at  $h = 4$ , above the random walk benchmark, using a single uncertainty measure and no additional predictors, such as yield curve, survey data, or financial variables. This suggests that output gap disagreement is not merely a contemporaneous symptom of downturns but carries some forward-looking information. A plausible interpretation is that methodological disagreement widens as the economy approaches a cyclical turning point, before the turning point becomes visible in the data, because different filtering and estimation methods respond at different speeds to emerging slack. If this interpretation is correct, the standard deviation of output gap estimates behaves as a genuine leading indicator, not only a coincident one, which strengthens the case for monitoring it as part of a broader recession early warning system.

The estimated recession probabilities from the best-performing specification are intuitive and track actual recession episodes closely. Figure 8 plots recession probabilities from the dynamic panel FE model at  $h = 1$ , with actual recession periods shown as shaded areas. Despite the short out-of-sample window, the model captures recessions well across countries. The dynamic specification produces smooth probability paths that remain low during expansions and rise sharply in the quarters preceding or coinciding with a recession, which is a desirable property for an early warning indicator. The model correctly identifies the COVID recession in all countries in the sample, and correctly assigns a low recession probability to Lithuania throughout that period, consistent with Lithuania not entering a technical recession in 2020. The model also captures post-COVID contractions in Austria, Estonia, Denmark, Finland, Latvia, Hungary, the Netherlands, Germany, Poland, and Slovakia. For countries that experienced a prolonged downturn following the Global Financial Crisis, the model picks up the tail end of that episode, since the out-of-sample period begins in 2012:Q1. A short out-of-sample window prevents a more comprehensive evaluation across a larger set of recession episodes, and this remains the principal constraint on inference.

A particularly sharp illustration of the model’s reliability comes from the COVID episode. The dynamic FE model at  $h = 1$ , shown in Figure 8, assigns near-zero recession probability to Lithuania throughout 2020, while correctly flagging all other countries as entering a recession. Lithuania did not meet the two-consecutive-quarters criterion during the pandemic and therefore does not appear as a recession in our binary indicator. The model, estimated entirely on pre-COVID data through the expanding window, recovers this distinction without any COVID-specific information. This is a meaningful out-of-sample validation, and it suggests the model responds to underlying economic conditions of particular countries.

We now turn to the comparison between the model estimated with the output gap standard deviation and the model estimated with the mean output gap (Model Averaged Output Gap – MAOG). In the panel setting, the model with the standard deviation does not outperform the model with the average, though it remains competitive in dynamic specifications. In country-by-country estimation, it does outperform the average-based model in a subset of countries, which we discuss below.

Figure 9 compares out-of-sample AUROC values for the panel probit model with the average output gap as the explanatory variable (bars) against the model with the standard deviation (dots). The model with the standard deviation performs at best as well as the model with the average, and in static specifications it performs considerably worse. In dynamic specifications the gap narrows, but the DeLong test rejects equality of AUROC at conventional significance levels for all model types except the dynamic FE model, where the null of equal AUROC cannot be rejected. The standard deviation-based model therefore does not outperform the average-based model in the panel setting, though the dynamic FE case is the closest to parity.

Turning to the country-by-country evidence, the model with the standard deviation reaches higher AUROC values than the average-based model in roughly one quarter to one third of countries, depending on the forecasting horizon.

Figure 10 presents out-of-sample AUROC values for the standard deviation model (blue dots) and the average model (red circles) at horizons  $h = 1$  to 4. Countries on the horizontal axis are ordered by the difference between the two AUROC values, so that countries where the standard deviation model performs relatively better appear on the left. Since this difference

varies across horizons, the country ordering differs across the four panels. Three horizontal reference lines are included in each panel: an AUROC of 0.5, and DeLong test p-value thresholds of 0.10 and 0.05, with the corresponding p-values shown as gray bars.

At  $h = 1$  the differences between the models are the biggest, and the standard deviation model reaches a higher AUROC in 10 out of 26 countries (38%). The DeLong test, however, fails to reject equality of AUROC in almost every country. At longer horizons the differences between the two models persist, but are smaller. At  $h = 2$  and  $h = 3$  the standard deviation model outperforms in 8 out of 26 countries (31%), while at  $h = 4$  it does so in 7 out of 26 countries (27%). Italy is a notable case: at  $h = 3$  and  $h = 4$  the standard deviation model substantially and statistically significantly outperforms the average model, with DeLong p-values well below 0.05 in both cases. This is, however, an isolated result rather than a systematic pattern.

The near-universal failure of the DeLong test to detect statistically significant differences between the two models warrants a brief discussion. The DeLong test estimates the variance of the AUROC difference using a structural components approach that is sensitive to the number of recession and non-recession observations. With at most three recessions per country in the out-of-sample window, the number of positive observations  $n_1$  is small, which inflates the standard error of the AUROC estimator and sharply reduces the power of the test. A difference of 0.10 to 0.13 in AUROC, which would be considered meaningful, can easily fall short of the threshold for statistical significance when  $n_1$  is below 15. The test results should therefore be read alongside the AUROC point estimates rather than in isolation. Where DeLong p-values are low, the evidence for a genuine difference is strong. Where they are high, the data are insufficient to resolve the question statistically, not necessarily evidence of equal performance.

Taken together, the results indicate that output gap model uncertainty, measured by the standard deviation of real-time estimates, is a meaningful recession predictor. Estimated recession probabilities are intuitive, track actual recession episodes well, and could serve as a practical early warning indicator. In the panel setting the standard deviation-based model does not systematically outperform the model that uses the average output gap, but in dynamic FE specifications the two are close to parity, and in certain countries the standard

deviation model offers a clear advantage. Given the parsimonious nature of the models and the short evaluation window, the overall forecasting performance of model uncertainty as a standalone predictor is encouraging.

## 5 Conclusion

This study investigates quasi-real-time model uncertainty of the model-averaged output gaps over the business cycle. We employ thirteen output gap methodologies to estimate the model-averaged output gap (MAOG) and its standard deviation as a measure of model uncertainty.

Our findings reveal that uncertainty exhibits high volatility over the business cycle, particularly escalating during recessions. Typically, slowdowns contribute to the widening of the standard deviation across output gap estimation methodologies. Such behavior is systematic rather than random.

To confirm such nonlinear systematic behavior, we use both linear and threshold dynamic panel models. The linear dynamic panel fixed-effects model demonstrates a strong negative effect of GDP growth rate on uncertainty, coupled with a high level of uncertainty persistence. Furthermore, the nonlinear threshold model reveals compelling evidence of nonlinearity. In models where GDP growth rate or Model Averaged Output Gap (MAOG) serve as threshold variables, a clear distinction arises between the regime below the threshold value, where the correlation between uncertainty and GDP growth is negative, statistically significant and robust, and the rest of the sample, where the effect of GDP growth on uncertainty is statistically not different from zero.

Furthermore, results suggest that the negative relationship between GDP growth and model uncertainty is typical for larger recessions, a not for mild slowdowns, which supports the plucking model of business cycles. Regarding absolute changes in GDP growth rates as a measure of the magnitude of the GDP shock, there is evidence that sample splitting improves model fit. However, the coefficients' signs in both regimes are negative, indicating that large absolute shocks can not explain the nonlinear nature of uncertainty in our estimates. Finally, we find that model uncertainty is quite persistent over the business cycle, but persistence is higher during periods of economic growth than during recessions. Taken together, the

evidence suggests that policy-making in real time is more complicated around the business cycle turning point.

Output gap model uncertainty is also a meaningful predictor of recessions. Estimated recession probabilities are intuitive, track actual recession episodes well, and could serve as a practical early warning indicator. AUROC values for shorter horizons in dynamic specifications are 0.87. They decrease at longer horizons, but the decay is rather slow (0.6 at the horizon  $h = 4$ ). This suggests that output gap disagreement is not merely a contemporaneous symptom of downturns but carries some forward-looking information. A plausible interpretation is that methodological disagreement widens as the economy approaches a cyclical turning point, before the turning point becomes visible in the data, because different filtering and estimation methods respond at different speeds to emerging slack. If this interpretation is correct, the standard deviation of output gap estimates behaves as a genuine leading indicator, not only a coincident one, which strengthens the case for monitoring it as part of a broader recession early warning system.

In the panel setting, the standard deviation-based model does not systematically outperform the model that uses the average output gap, but in the dynamic fixed-effects specification, the two are statistically indistinguishable, and in certain countries, the standard deviation model offers a clear advantage. The results suggest that output gap model uncertainty varies systematically over the business cycle and contains valuable information for recession forecasting.

A practical feature of the proposed forecasting models is that all variables are available in real time. The standard deviation of output gap estimates can be computed as soon as a new vintage of GDP data becomes available and the estimates are updated across methods, with no requirement for data subject to substantial revision or publication lags. This distinguishes the model from many forecasting approaches that rely on financial market variables, survey indicators, or yield curve measures, which may not be available at the same frequency or with the same timeliness across all EU member states. The model is therefore operationally viable as a short-horizon early warning tool for national authorities or international institutions that monitor recession risk in real time.

In essence, our study underscores the importance of understanding and addressing un-

certainty in output gap estimation methodologies, as it significantly influences the decision-making processes in economic policy formulation. Further research is warranted to delve deeper into the dynamics of uncertainty and its implications for economic policymaking, particularly in mitigating the adverse effects of recessions and enhancing overall economic stability and resilience.

Our results also support the significance of nonlinear, multiple regime treatment of the analysis of the business cycle in the fashion of [Friedman \(1993\)](#); [Bradley and Jansen \(2017\)](#) and [Huang et al. \(2021\)](#). One of the possible venues of research is the exploration of model uncertainty of MAOG estimates using nonlinear output gap estimation techniques such as Markov switching methodology and/or BBU, BBV and/or BBD nonlinear gaps ([Morley and Panovska, 2020](#); [Arčabić et al., 2024](#)).

In the realm of economic theory, especially when it comes to welfare functions in the design of the optimal policy, it would be interesting to see how can we redesign the welfare function to take uncertainty into account. Minimization of the welfare loss function is usually a negative function of output volatility in an almost deterministic way, and uncertainties related to real-time data availability and/or model choice are usually not taken into account (see [Galí \(2015\)](#) for example).

There are several limitations to this paper. Firstly, it does not encompass numerous other output gap estimation methodologies. Secondly, the addition of genuine instead of quasi-real-time samples (vintage data before revisions) could enhance the applicability of the results for policymakers even further. Thirdly, employing bootstrapping for confidence intervals could provide a deeper understanding of the nonlinear nature of the data.

## References

- Arčabić, V. and Olson, E. (2019). Estimates of okun’s law using a new output gap measure. *Economics Bulletin*, 39(2):929–936.
- Arčabić, V., Panovska, I., and Tica, J. (2024). Business cycle synchronization and asymmetry in the european union. *Economic modelling*, 139:106811.
- Barkema, J., Gudmundsson, T., and Mrkaic, M. M. (2020). What do we talk about when we talk about output gaps? *IMF Working Papers*, 2020/259.
- Baxter, M. and King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, 81(4):575–593.
- Bayoumi, T. and Eichengreen, B. (1992). Shocking aspects of european monetary unification. NBER Working Paper 3949, National Bureau of Economic Research.
- Beveridge, S. and Nelson, C. R. (1981). A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the ‘Business Cycle’. *Journal of Monetary Economics*, 7(2):151–174.
- Blanchard, O. J. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79(4):655–673.
- Bradley, M. D. and Jansen, D. W. (2017). Nonlinear Evidence on the Existence of Jobless Recoveries. *Studies in Nonlinear Dynamics & Econometrics*, 22(1):20160081.
- Cattelan, G. and Nandwa, M. B. (2024). *Output Gap Uncertainty and Fiscal Policy Adjustment in Real-Time in Emerging Economies*. International Monetary Fund.
- Chalaux, T. and Guillemette, Y. (2019). The oecd potential output estimation methodology. Technical Report 1563, OECD Economics Department Working Papers.
- Chalaux, T., Turner, D., and Cassimon, S. (2025). Harnessing the wisdom of crowds to assess recession risks in oecd countries. *OECD Economics Department Working Papers*.
- Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21(1):520–533.
- Christiano, L. J. and Fitzgerald, T. J. (1999). The band pass filter. Working Paper 7257, National Bureau of Economic Research.
- Clark, P. K. (1987). The cyclical component of us economic activity. *The Quarterly Journal of Economics*, 102(4):797–814.
- De Masi, P. (1997). *IMF Estimates of Potential Output: Theory and Practice*. International Monetary Fund. <https://www.imf.org/en/Publications/WP/Issues/2016/12/30/IMF-Estimates-of-Potential-Output-Theory-and-Practice-2451>.

- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845.
- Donayre, L. and Panovska, I. (2021). Recession-specific recoveries: L’s, u’s and everything in between. *Economics Letters*, 209:110145.
- Dupraz, S., Nakamura, E., and Steinsson, J. (2025). A plucking model of business cycles. *Journal of Monetary Economics*, 152:103766.
- Enders, W., Falk, B., and Siklos, P. L. (2007). A threshold model of real us gdp and the problem of constructing confidence intervals in tar models. *Studies in Nonlinear Dynamics and Econometrics*, *Forthcoming*.
- EU Independent Fiscal Institutions (2020). A practitioner’s guide to potential output and the output gap. Available at: <https://www.euifis.eu/publications/15>.
- European Commission (2025). Economic forecasts. [https://economy-finance.ec.europa.eu/economic-forecast-and-surveys/economic-forecasts\\_en](https://economy-finance.ec.europa.eu/economic-forecast-and-surveys/economic-forecasts_en).
- Eurostat (2025a). Employment and activity by sex and age - quarterly data. [https://ec.europa.eu/eurostat/databrowser/view/LFSI\\_EMP\\_Q\\_\\_custom\\_7378257/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/LFSI_EMP_Q__custom_7378257/default/table?lang=en).
- Eurostat (2025b). Employment and activity by sex and age (1989-2020) - quarterly data. [https://ec.europa.eu/eurostat/databrowser/view/LFSI\\_EMP\\_Q\\_H/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/LFSI_EMP_Q_H/default/table?lang=en).
- Eurostat (2025c). Gdp and main components (output, expenditure and income). Data retrieved from [https://ec.europa.eu/eurostat/databrowser/view/NAMQ\\_10\\_GDP/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/NAMQ_10_GDP/default/table?lang=en).
- Eurostat (2025d). Gross domestic product (gdp) and main components (output, expenditure and income). [https://ec.europa.eu/eurostat/databrowser/view/namq\\_10\\_gdp\\_custom\\_20857148/default/table](https://ec.europa.eu/eurostat/databrowser/view/namq_10_gdp_custom_20857148/default/table).
- Eurostat (2025e). Unemployment by sex and age (1992-2020) - quarterly data. Data retrieved from [https://ec.europa.eu/eurostat/databrowser/view/UNE\\_RT\\_Q\\_H/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/UNE_RT_Q_H/default/table?lang=en).
- Eurostat (2025f). Unemployment rates by sex, age and citizenship (%). Data retrieved from [https://ec.europa.eu/eurostat/databrowser/view/LFSQ\\_URGAN/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/LFSQ_URGAN/default/table?lang=en).
- Friedman, M. (1993). The “Plucking Model” of Business Fluctuations Revisited. *Economic Inquiry*, 31(2):171–177.
- Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Economics Books. Princeton University Press.

- Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal*, 7(1):98–119.
- Guérin, P., Maurin, L., and Mohr, M. (2015). Trend-cycle decomposition of output and euro area inflation forecasts: A real-time approach based on model combination. *Macroeconomic Dynamics*, 19(2):363–393.
- Hagelund, K., Hansen, F., and Robstad, O. (2018). Model estimates of the output gap. Research Report 4/2018, Staff Memo.
- Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.
- Hamilton, J. D. (2018). Why You Should Never Use the Hodrick-Prescott Filter. *The Review of Economics and Statistics*, 100(5):831–843.
- Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.
- Harvey, A. C. (1985). Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics*, 3(3):216–227.
- Havik, K., Morrow, K. M., Orlandi, F., Planas, C., Raciborski, R., Roeger, W., Rossi, A., Thum-Thysen, A., and Vandermeulen, V. (2014). The production function methodology for calculating potential growth rates & output gaps. Technical Report 535, European Economy - Economic Papers 2008 - 2015.
- Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. pages 179–195.
- Hodrick, R. J. and Prescott, E. (1981). Post-war u.s. business cycles: An empirical investigation. Technical Report 451, Discussion Papers.
- Huang, Y.-F., Luo, S., and Startz, R. (2021). Are Recoveries All the Same: GDP and TFP? *Oxford Bulletin of Economics and Statistics*, 85(5):1111–1129.
- IMF (2025). World economic outlook databases. <https://www.imf.org/en/Publications/SPROLLs/world-economic-outlook-databases#sort=%40imfdate%20descending>.
- Kamber, G., Morley, J., and Wong, B. (2018). Intuitive and reliable estimates of the output gap from a beveridge-nelson filter. *Review of Economics and Statistics*, 100(3):550–566.
- Kamber, G., Morley, J., and Wong, B. (2025). Trend-cycle decomposition in the presence of large shocks. *Journal of Economic Dynamics and Control*, 173:105066.
- Kim, C.-J., Morley, J., and Piger, J. (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics*, 20(2):291–309.

- Kim, C.-J. and Nelson, C. R. (1999). Friedman’s plucking model of business fluctuations: tests and estimates of permanent and transitory components. *Journal of Money, Credit and Banking*, pages 317–334.
- Kumar, L. (2026). The output gap: Method choice, data revisions, and policy implications. Technical Report 127829, MPRA.
- Laxton, D. and Tetlow, R. (1992). A simple multivariate filter for the measurement of potential output. Technical Report 59, Bank of Canada.
- Massmann, M. and Mitchell, J. (2005). Reconsidering the evidence. *Journal of Business Cycle Measurement and Analysis*, 2004(3):275–307.
- Mazzi, G. L., Ozyildirim, A., and Mitchell, J. (2017). Alternative detrending methods. In *Handbook on Cyclical Composite Indicators for Business Cycle Analysis*. Publications Office of the European Union and United Nations, Luxembourg and New York.
- Morley, J. (2014). Measuring economic slack: A forecast-based approach with applications to economies in asia and the pacific. BIS Working Paper 451, Bank for International Settlements.
- Morley, J. and Panovska, I. B. (2020). Is business cycle asymmetry intrinsic in industrialized economies? *Macroeconomic Dynamics*, 24(6):1403–1436.
- Morley, J. and Piger, J. (2012). The asymmetric business cycle. *The Review of Economics and Statistics*, 94(1):208–221.
- Morley, J. C., Nelson, C. R., and Zivot, E. (2003). Why are the beveridge-nelson and unobserved-components decompositions of gdp so different? *The Review of Economics and Statistics*, 85(2):235–243.
- Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.
- OECD (2025a). Economic projections, oecd economic outlook. Data retrieved from <https://stats.oecd.org/index.aspx?DataSetCode=EO>.
- OECD (2025b). Quarterly national accounts: Quarterly real gdp growth. Data retrieved from <https://stats.oecd.org/index.aspx?queryid=350>.
- OECD (2025c). Unemployment rate. Data retrieved from <https://data.oecd.org/unemp/unemployment-rate.htm>.
- Office for National Statistics (2025). Gdp deflator. <https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyu>.
- Onatski, A. and Williams, N. (2003). Modeling Model Uncertainty. *Journal of the European Economic Association*, 1(5):1087–1122.

- Orphanides, A. and Norden, S. V. (2000). The reliability of output gap estimates in real time. In *Econometric Society World Congress 2000 Contributed Papers*, volume 0768. Econometric Society.
- Orphanides, A. and van Norden, S. (2002). The unreliability of output-gap estimates in real time. *Review of economics and statistics*, 84(4):569–583.
- Perron, P. and Wada, T. (2009). Let’s take a break: Trends and cycles in us real gdp. *Journal of Monetary Economics*, 56(6):749–765.
- Quast, J. and Wolters, M. H. (2022). Reliable real-time output gap estimates based on a modified hamilton filter. *Journal of Business & Economic Statistics*, 40(1):152–168.
- Segal, G. (2017). To respond or not to respond: Measures of the output gap in theory and in practice. *International Journal of Central Banking*, 13(2):73–120.
- Sinclair, T. M. (2009). The relationships between permanent and transitory movements in us output and the unemployment rate. *Journal of Money, Credit and Banking*, 41(2-3):529–542.
- Woitek, U. (1998). A note on the baxter-king filter. Working paper, Business School - Economics, University of Glasgow.
- Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, non-linear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1):39–54.

# Tables

Table 1: Linear model with quasi-real-time output gap estimates

|               | (1)<br>ReSD          | (2)<br>ReSD          | (3)<br>ReSD          | (4)<br>ReSD          | (5)<br>ReSD          | (6)<br>ReSD          |
|---------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy            | -0.130***<br>(0.000) | -0.137***<br>(0.000) | -0.076***<br>(0.000) | -0.087***<br>(0.006) | -0.088***<br>(0.000) | -0.032***<br>(0.000) |
| lag1dependent |                      |                      | 0.586***<br>(0.000)  |                      |                      | 0.801***<br>(0.000)  |
| lag4dependent |                      | 0.513***<br>(0.000)  | 0.163***<br>(0.006)  |                      | 0.575***<br>(0.000)  | 0.042<br>(0.168)     |
| _cons         | 2.453***<br>(0.000)  | 1.353***<br>(0.000)  | 0.708***<br>(0.000)  | 2.178***<br>(0.000)  | 0.723***<br>(0.000)  | 0.028<br>(0.851)     |
| $N$           | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| countryFE     | Yes                  | Yes                  | Yes                  | Yes                  | Yes                  | Yes                  |
| timeFE        |                      |                      |                      | Yes                  | Yes                  | Yes                  |
| rmse          | 1.038                | 0.858                | 0.695                | 0.926                | 0.750                | 0.504                |
| aic           | 6966.339             | 5791.866             | 4833.338             | 6374.422             | 5133.295             | 3314.283             |
| bic           | 6972.119             | 5803.336             | 4850.544             | 6518.919             | 5276.680             | 3457.669             |

**Notes:**  $p$ -values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ .  $dy$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent is  $\delta_{it-1}$  and lag4dependent is  $\delta_{it-4}$ .

Table 2: GDP growth rate as the threshold variable, nonlinear model with quasi-real-time output gap estimates

|                  | (1)<br>ReSD, d=0     | (2)<br>ReSD, d=0     | (3)<br>ReSD, d=0     | (4)<br>ReSD, d=1     | (5)<br>ReSD, d=1     | (6)<br>ReSD, d=1     |
|------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy_1             | -0.303***<br>(0.000) | -0.294***<br>(0.000) | -0.201***<br>(0.000) | -0.185***<br>(0.000) | -0.185***<br>(0.000) | -0.131***<br>(0.000) |
| lag4dependent_1  |                      | 0.429***<br>(0.000)  | 0.173<br>(0.135)     |                      | 0.558***<br>(0.000)  | 0.236**<br>(0.014)   |
| lag1dependent_1  |                      |                      | 0.507***<br>(0.000)  |                      |                      | 0.533***<br>(0.000)  |
| dy_2             | 0.049<br>(0.301)     | -0.003<br>(0.893)    | 0.000<br>(0.972)     | -0.016<br>(0.640)    | -0.022<br>(0.219)    | -0.005<br>(0.414)    |
| lag4dependent_2  |                      | 0.567***<br>(0.000)  | -0.032**<br>(0.041)  |                      | 0.581***<br>(0.000)  | -0.070**<br>(0.014)  |
| lag1dependent_2  |                      |                      | 0.895***<br>(0.000)  |                      |                      | 0.941***<br>(0.000)  |
| _cons            | 1.800***<br>(0.000)  | 0.464**<br>(0.021)   | -0.113<br>(0.455)    | 1.992***<br>(0.000)  | 0.496***<br>(0.008)  | -0.112<br>(0.446)    |
| <i>N</i>         | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| <i>Tr_value</i>  | 0.029                | -0.131               | -0.993               | -1.015               | -1.015               | -1.015               |
| <i>Wald_pval</i> | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                |
| <i>rmse</i>      | 0.856                | 0.690                | 0.466                | 0.892                | 0.708                | 0.471                |
| <i>aic</i>       | 5999.656             | 4752.722             | 2956.120             | 6194.156             | 4871.807             | 2998.394             |
| <i>bic</i>       | 6144.153             | 4896.108             | 3099.506             | 6338.653             | 5015.193             | 3141.780             |

**Notes:** *p*-values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ . In regime 1, below threshold value  $\theta$ ,  $dy_1$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_1 is  $\delta_{it-1}$  and lag4dependent\_1 is  $\delta_{it-4}$ . In regime 2, above threshold value  $\theta$ ,  $dy_2$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_2 is  $\delta_{it-1}$  and lag4dependent\_2 is  $\delta_{it-4}$ . *Tr\_value* presents estimated threshold values  $\theta$ . *Wald\_pval* is the *p*-value for the Wald test with  $H_0 : dy_1 = dy_2$ . Threshold variable  $q_{it-d}$  is  $\Delta y_{it-d}$  and delay parameter *d* is lag of the threshold variable. Threshold grid search results are presented in Figure 4.

Table 3: GDP growth rate as the threshold variable, nonlinear model with quasi-real-time output gap estimates, 5th to 95th percentile grid search

|                 | (1)<br>ReSD, d=0     | (2)<br>ReSD, d=0     | (3)<br>ReSD, d=0     | (4)<br>ReSD, d=1     | (5)<br>ReSD, d=1     | (6)<br>ReSD, d=1     |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy_1            | -0.357***<br>(0.000) | -0.374***<br>(0.000) | -0.305***<br>(0.000) | -0.211***<br>(0.000) | -0.211***<br>(0.000) | -0.152***<br>(0.000) |
| lag4dependent_1 |                      | 0.257***<br>(0.003)  | 0.074<br>(0.429)     |                      | 0.487***<br>(0.000)  | 0.225**<br>(0.031)   |
| lag1dependent_1 |                      |                      | 0.400***<br>(0.000)  |                      |                      | 0.476***<br>(0.000)  |
| dy_2            | 0.053<br>(0.142)     | 0.000<br>(0.980)     | -0.006<br>(0.305)    | -0.032<br>(0.399)    | -0.025<br>(0.145)    | -0.023***<br>(0.002) |
| lag4dependent_2 |                      | 0.500***<br>(0.000)  | 0.017<br>(0.453)     |                      | 0.505***<br>(0.000)  | -0.025<br>(0.373)    |
| lag1dependent_2 |                      |                      | 0.812***<br>(0.000)  |                      |                      | 0.878***<br>(0.000)  |
| _cons           | 1.685***<br>(0.000)  | 0.617***<br>(0.001)  | 0.030<br>(0.810)     | 1.928***<br>(0.000)  | 0.688***<br>(0.000)  | 0.040<br>(0.741)     |
| <i>N</i>        | 2080                 | 1976                 | 1976                 | 2080                 | 1976                 | 1976                 |
| Tr_value        | 0.032                | -1.169               | -3.604               | 3.772                | -0.191               | -1.758               |
| Wald_pval       | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                |
| rmse            | 0.800                | 0.683                | 0.493                | 0.855                | 0.725                | 0.521                |
| aic             | 4939.780             | 4068.903             | 2778.546             | 5217.765             | 4304.899             | 2998.209             |
| bic             | 5080.783             | 4208.624             | 2918.267             | 5358.768             | 4444.620             | 3137.930             |

**Notes:** *p*-values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ . In regime 1, below threshold value  $\theta$ ,  $dy_1$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_1 is  $\delta_{it-1}$  and lag4dependent\_1 is  $\delta_{it-4}$ . In regime 2, above threshold value  $\theta$ ,  $dy_2$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_2 is  $\delta_{it-1}$  and lag4dependent\_2 is  $\delta_{it-4}$ . *Tr\_value* presents estimated threshold values  $\theta$ . *Wald\_pval* is the *p*-value for the Wald test with  $H_0 : dy_1 = dy_2$ . Threshold variable  $q_{it-d}$  is  $\Delta y_{it-d}$  and delay parameter *d* is lag of the threshold variable. Threshold grid search results are presented in Figure B.1.

Table 4: GDP growth rate as the threshold variable, nonlinear model with quasi-real-time output gap estimates, exogenous threshold model

|                 | (1)<br>ReSD, d=0     | (2)<br>ReSD, d=0     | (3)<br>ReSD, d=0     | (4)<br>ReSD, d=1     | (5)<br>ReSD, d=1     | (6)<br>ReSD, d=1     |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy_1            | -0.303***<br>(0.000) | -0.292***<br>(0.000) | -0.170***<br>(0.000) | -0.183***<br>(0.000) | -0.183***<br>(0.000) | -0.116***<br>(0.000) |
| lag4dependent_1 |                      | 0.432***<br>(0.000)  | 0.173**<br>(0.043)   |                      | 0.560***<br>(0.000)  | 0.204**<br>(0.021)   |
| lag1dependent_1 |                      |                      | 0.555***<br>(0.000)  |                      |                      | 0.586***<br>(0.000)  |
| dy_2            | 0.049<br>(0.301)     | -0.004<br>(0.875)    | 0.000<br>(0.980)     | -0.017<br>(0.628)    | -0.024<br>(0.212)    | -0.002<br>(0.765)    |
| lag4dependent_2 |                      | 0.566***<br>(0.000)  | -0.055**<br>(0.017)  |                      | 0.582***<br>(0.000)  | -0.075***<br>(0.009) |
| lag1dependent_2 |                      |                      | 0.904***<br>(0.000)  |                      |                      | 0.932***<br>(0.000)  |
| _cons           | 1.800***<br>(0.000)  | 0.468**<br>(0.020)   | -0.076<br>(0.616)    | 2.003***<br>(0.000)  | 0.503***<br>(0.008)  | -0.085<br>(0.556)    |
| <i>N</i>        | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| Tr_value        | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                |
| Wald_pval       | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                |
| rmse            | 0.856                | 0.690                | 0.472                | 0.894                | 0.711                | 0.476                |
| aic             | 5999.655             | 4754.404             | 3007.222             | 6207.275             | 4887.539             | 3049.474             |
| bic             | 6144.152             | 4897.790             | 3150.608             | 6351.772             | 5030.925             | 3192.860             |

**Notes:** *p*-values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ . In regime 1, below threshold value  $\theta$ ,  $dy_1$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_1 is  $\delta_{it-1}$  and lag4dependent\_1 is  $\delta_{it-4}$ . In regime 2, above threshold value  $\theta$ ,  $dy_2$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_2 is  $\delta_{it-1}$  and lag4dependent\_2 is  $\delta_{it-4}$ . *Tr\_value* is exogenously set to be equal to zero  $\theta = 0$ . *Wald\_pval* is the *p*-value for the Wald test with  $H_0 : dy_1 = dy_2$ . Threshold variable  $q_{it-d}$  is  $\Delta y_{it}$  and delay parameter *d* is lag of the threshold variable.

# Figures

Figure 1: Quasi-real-time output gaps

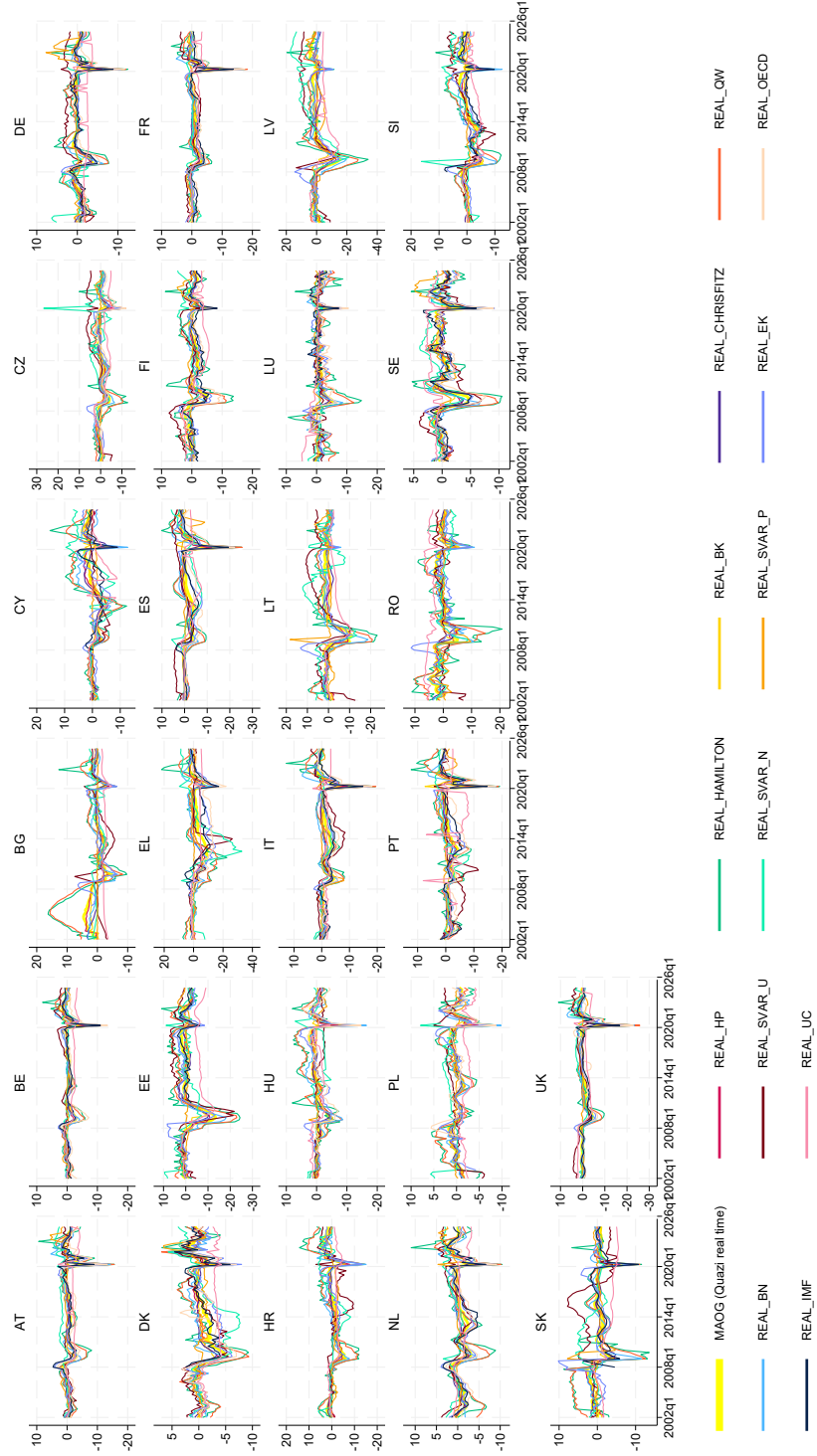


Figure 2: Standard deviation for the quasi-real-time output gap estimates

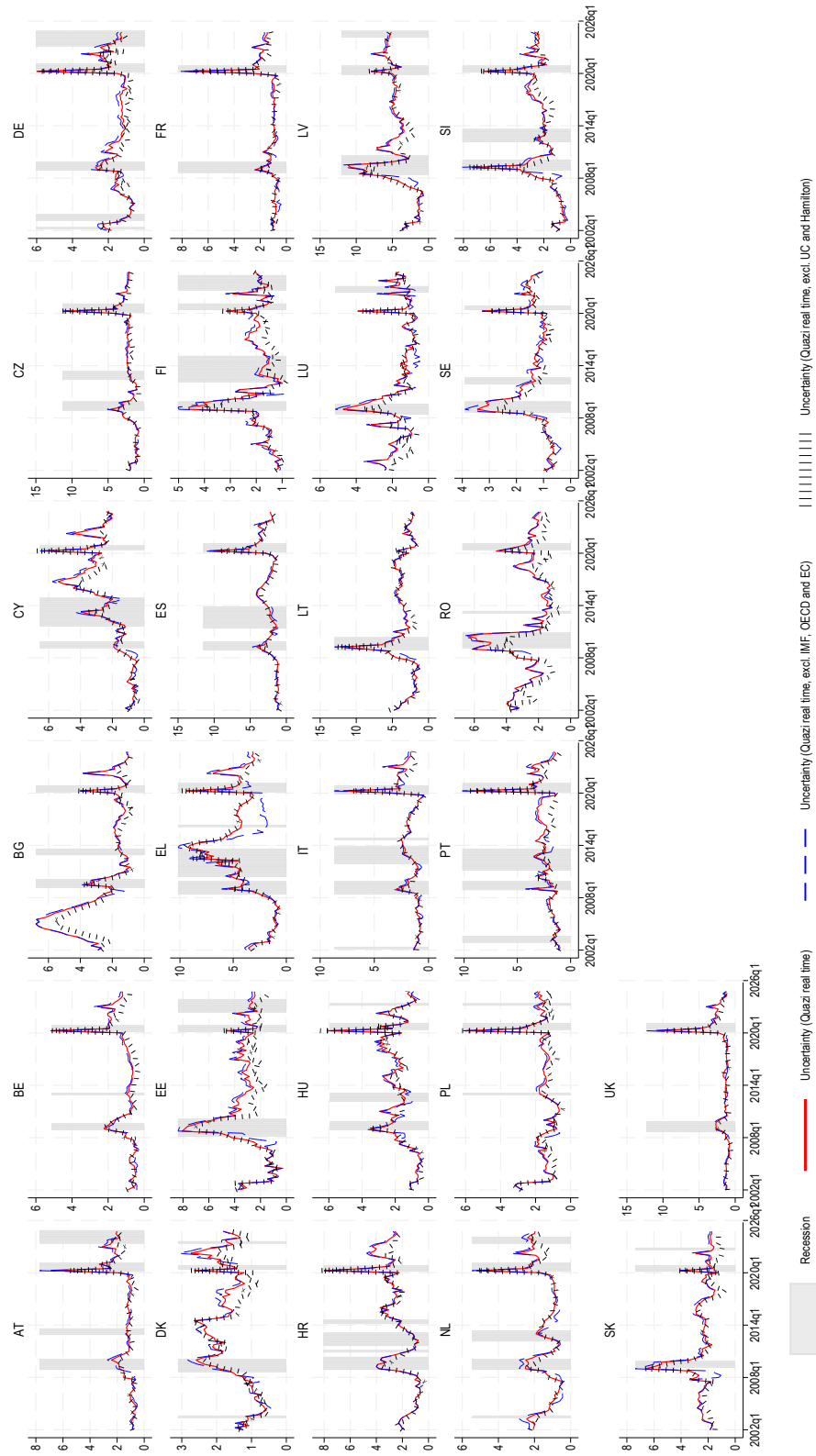


Figure 3: Standard deviation decomposition by output gap methods – Quasi real time gaps

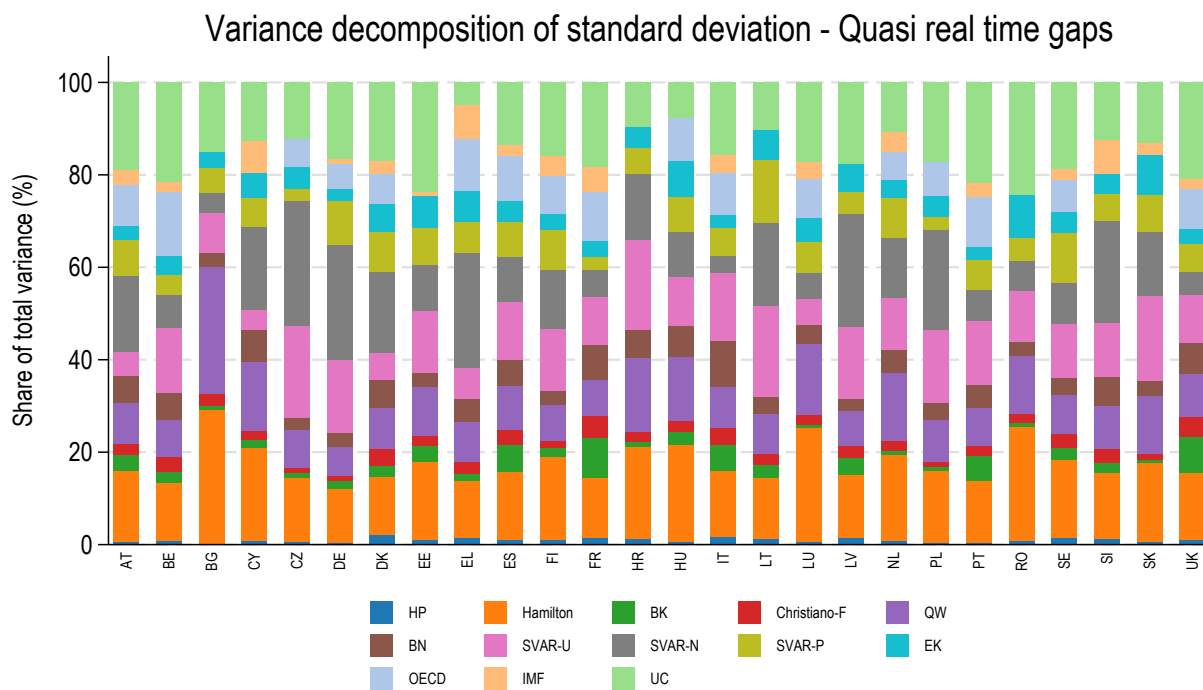
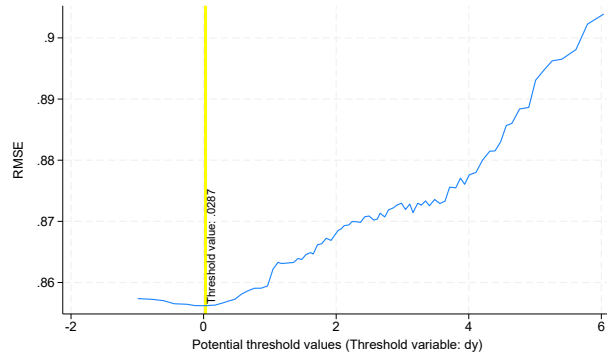
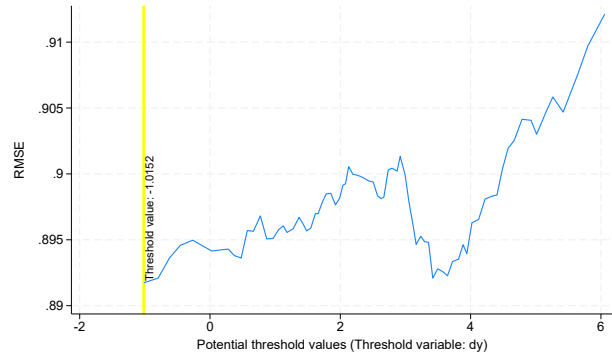


Figure 4: Estimated RMSE vs potential threshold values (GDP growth rate as threshold variable)

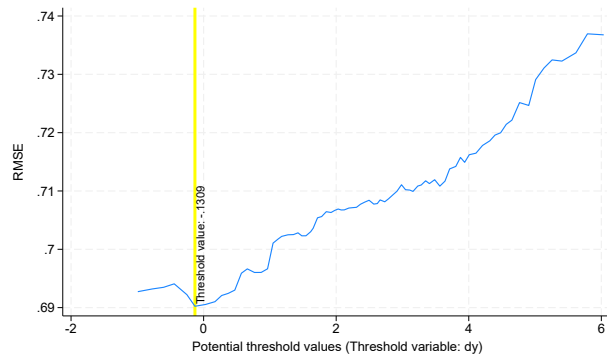
(a) Model 1: RMSE vs threshold values (delay=0)



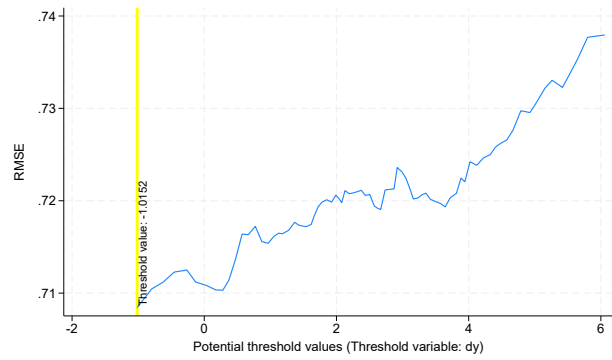
(b) Model 4: RMSE vs threshold values (delay=1)



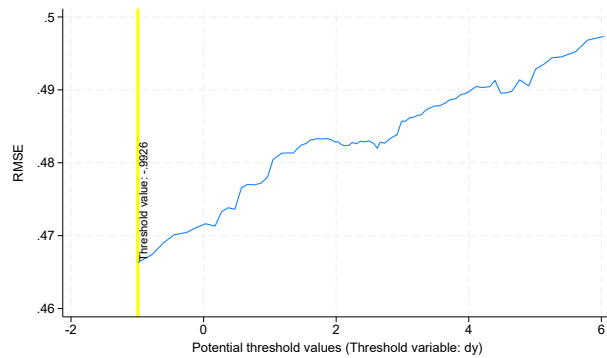
(c) Model 2: RMSE vs threshold values (delay=0)



(d) Model 5: RMSE vs threshold values (delay=1)



(e) Model 3: RMSE vs threshold values (delay=0)



(f) Model 6: RMSE vs threshold values (delay=1)

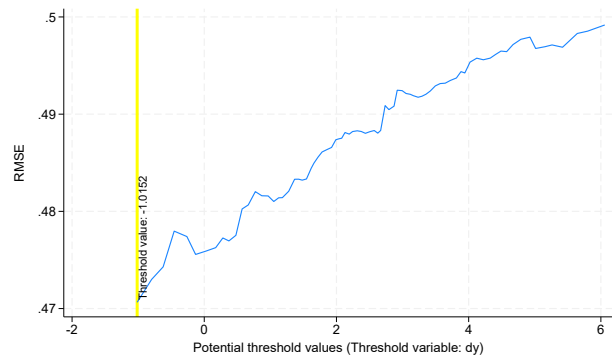


Figure 5: Uncertainty vs. GDP growth rate in two regimess and grid search span

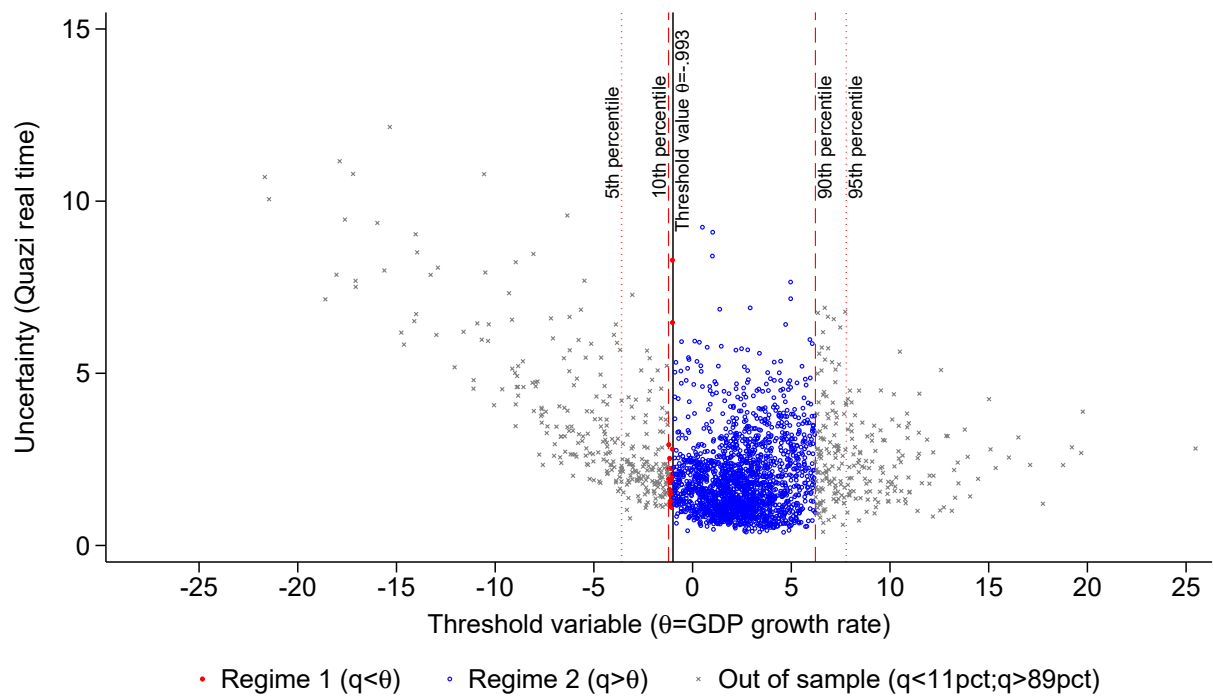


Figure 6: Uncertainty vs. GDP growth rate in two regime

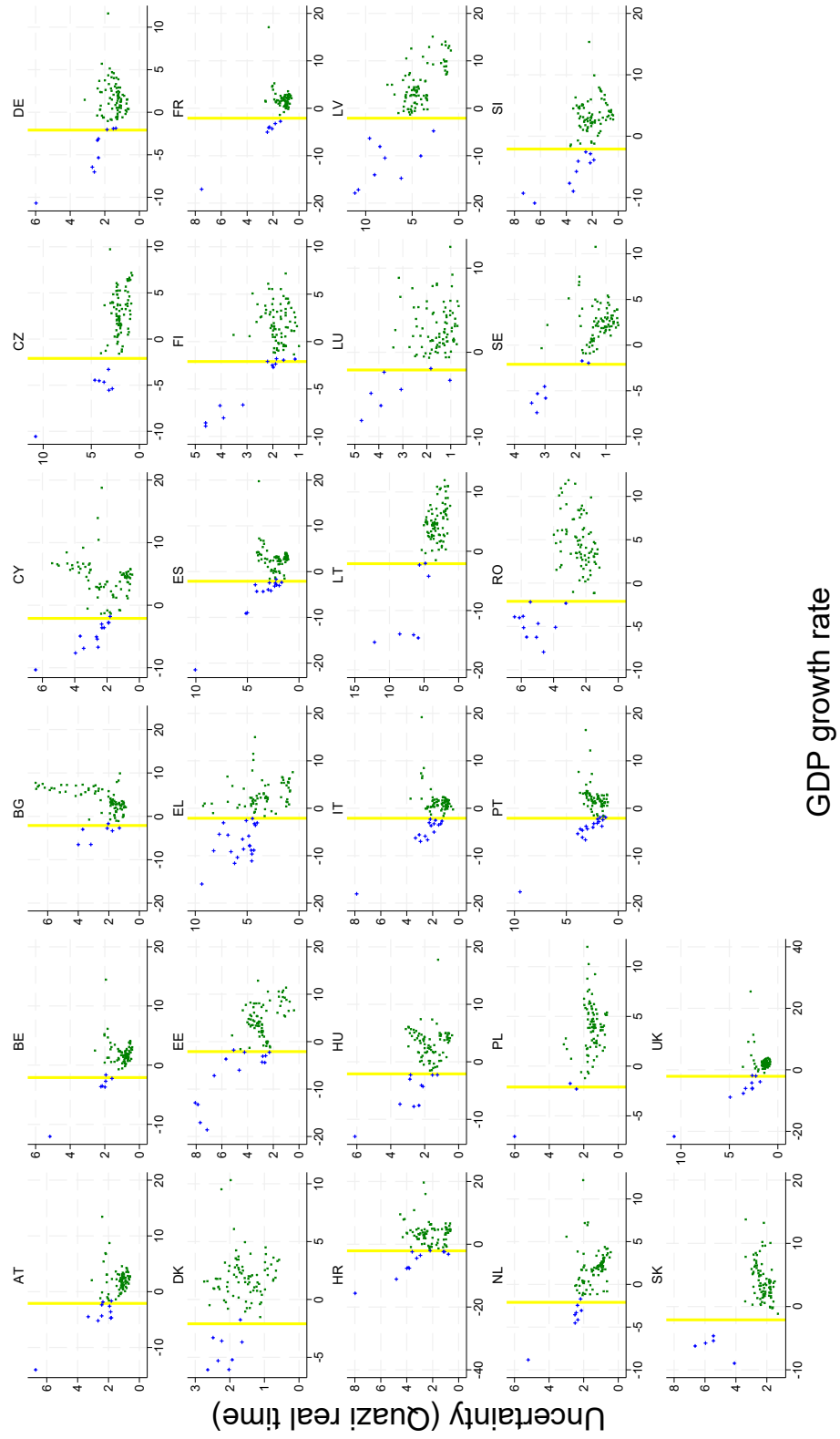


Figure 7: Panel probit AUROC by model type, dynamics, and lags

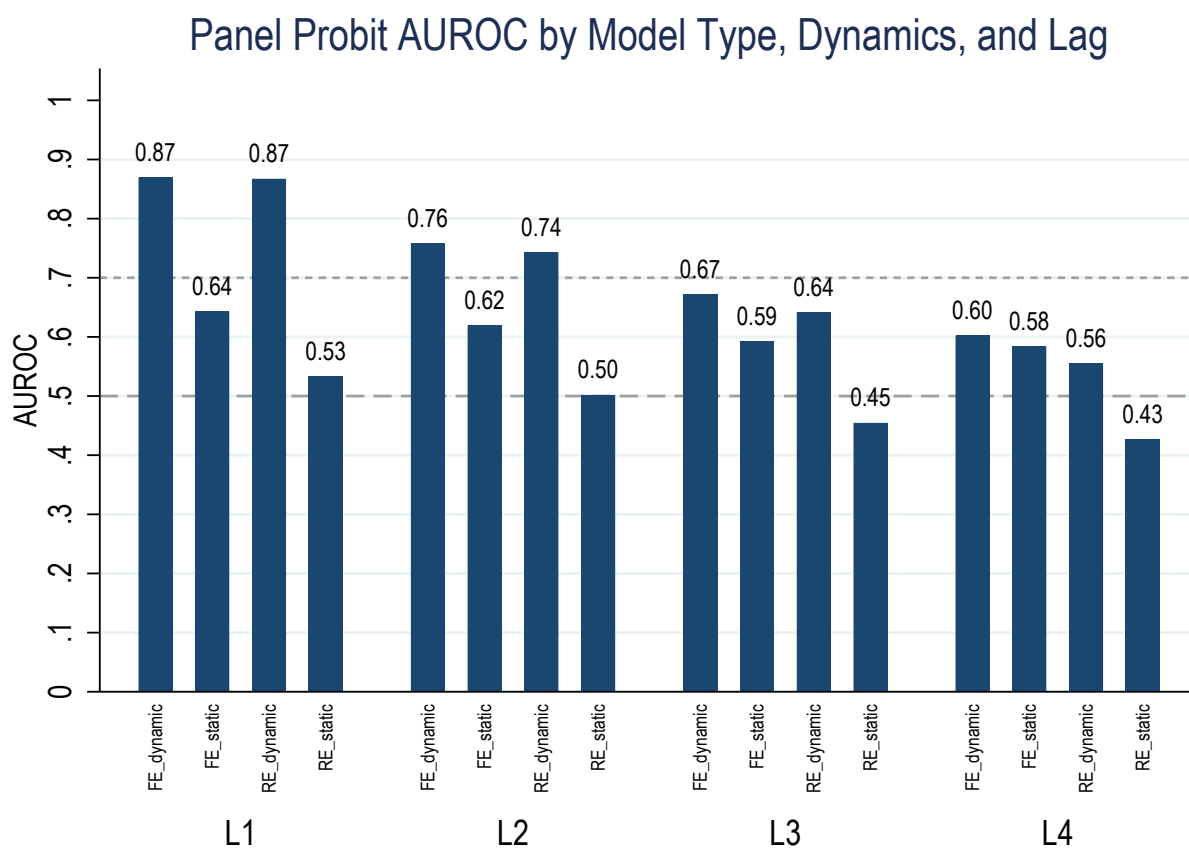


Figure 8: Recessions and probabilities from dynamic panel probit model with FE (OOS with 1 lag)

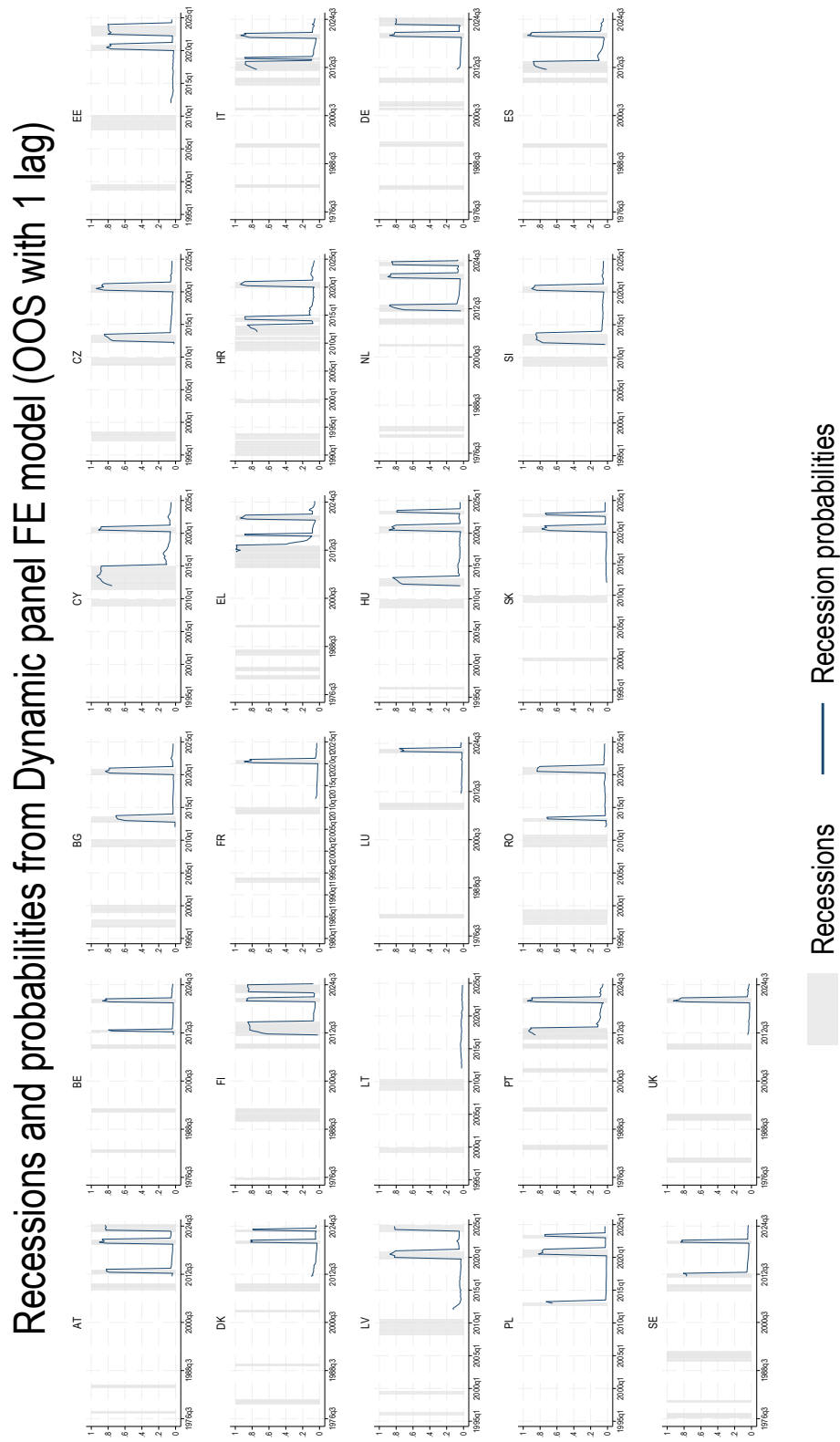


Figure 9: Panel probit AUROC - Average vs. Standard deviation

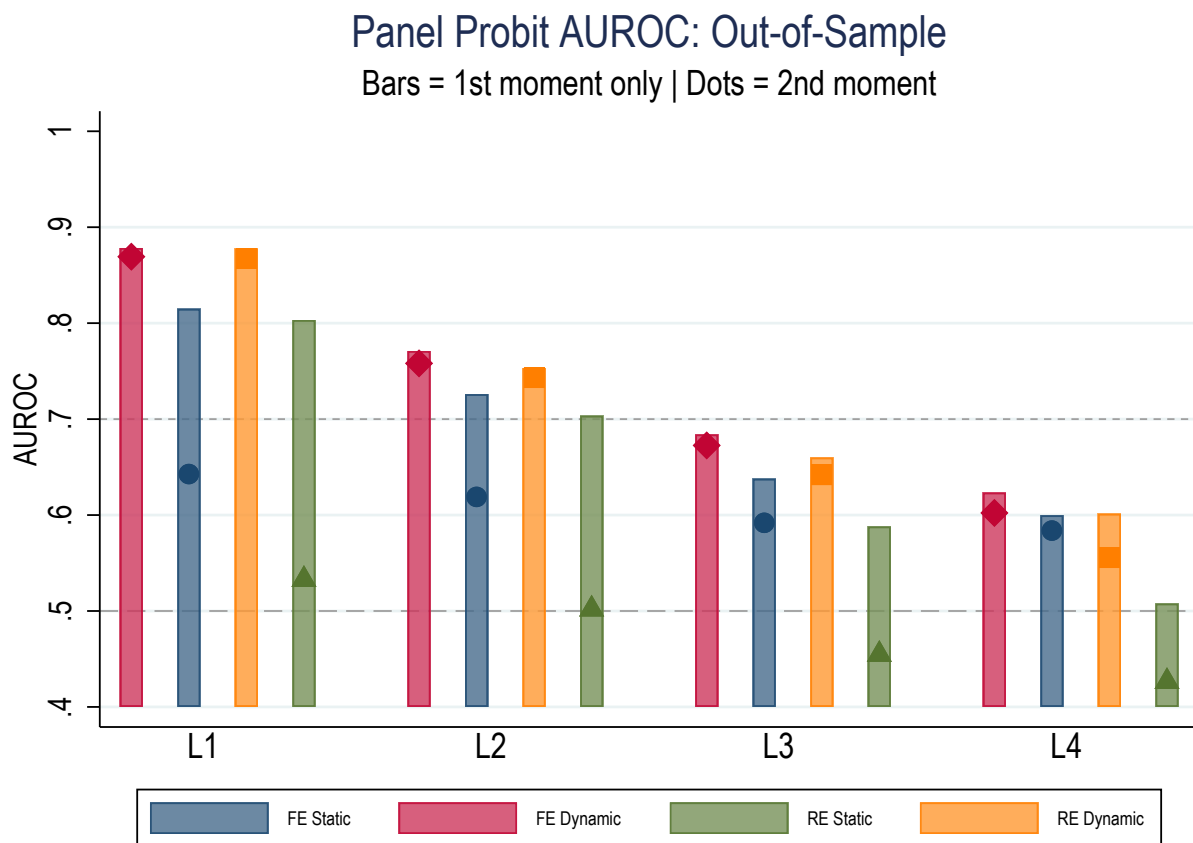
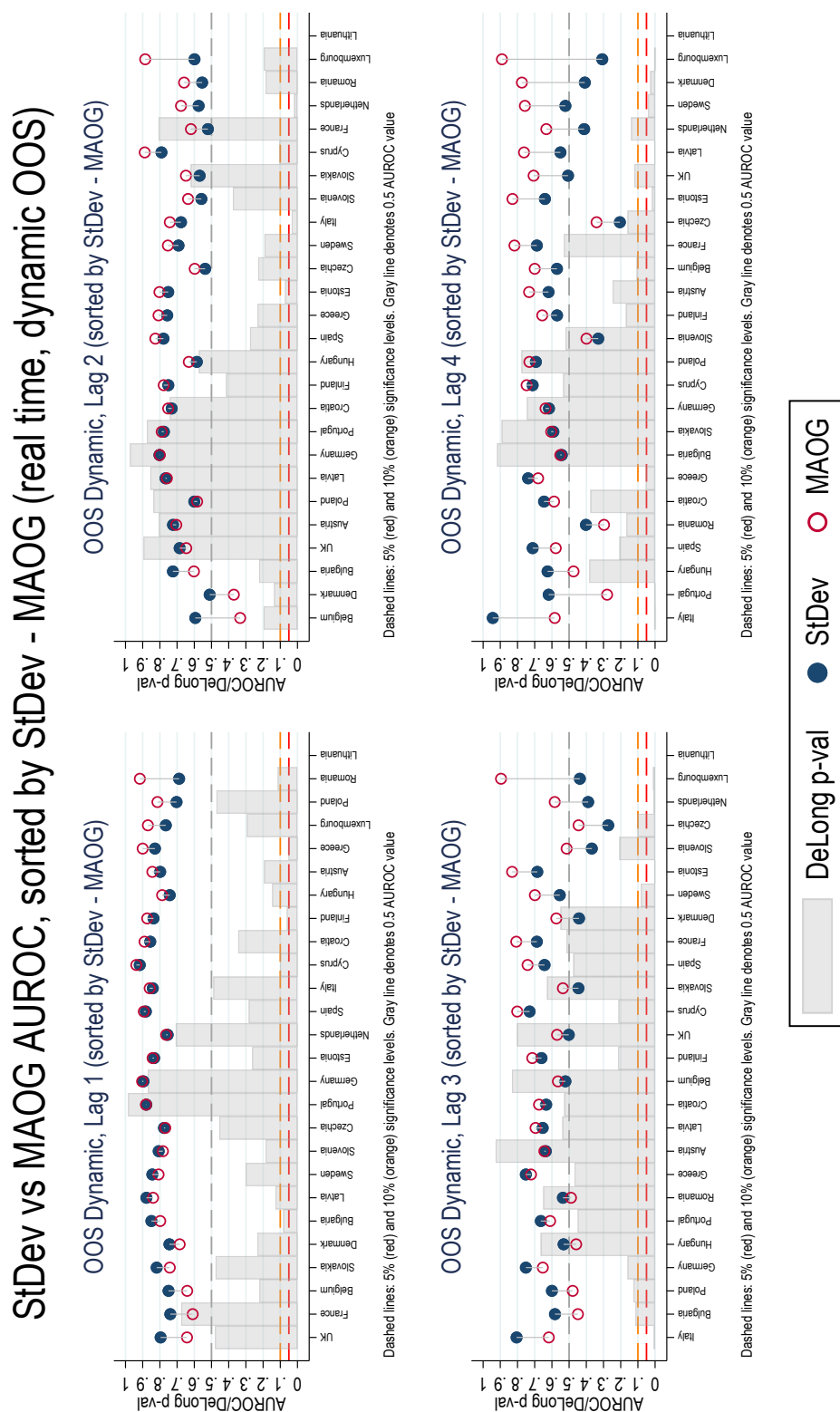


Figure 10: Country-by-country AUROC - Average vs. Standard deviation



## Appendix A: A detailed methodology description and data figures

The HP filter assumes a smooth trend when used to extract business cycles and minimizes the following sum with respect to  $\tau_t$ :

$$\min_{\tau_t} \left( \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right). \quad (8)$$

where  $y_t$  is the output,  $\tau_t$  is the output trend, and the difference between output and potential output  $c_t = y_t - \tau_t$  is the output gap. We use  $\lambda = 1,600$  as the smoothing parameter, which is common for quarterly data.

The second method is the linear [Hamilton \(2018\)](#) filter. This filter is better than the HP filter in the case of spurious cycles and due to its real-time properties. It is based on a simple OLS regression approach:

$$y_{t+8} = \alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1} + \alpha_3 y_{t-2} + \alpha_4 y_{t-3} + c_t, \quad (9)$$

where the residual  $c_t$  represents the output gap.

[Quast and Wolters \(2022\)](#) point out that the Hamilton filter has the drawback of not covering typical business-cycle frequencies evenly, as it mutes short cycles and amplifies medium-length cycles. Moreover, since the Hamilton filter's trend component contains high-frequency noise, it is harder to claim that the estimated trend represents potential GDP. They propose a modification of the Hamilton filter that better covers typical business-cycle frequencies (6-32 quarters) and provides a smoother estimate of the trend component. Instead of using a fixed 8-quarter horizon, they use an equally weighted average of forecast errors based on 4 to 12 quarter ahead projections to estimate the output gap  $c_t$ :

$$\begin{aligned} c_t &= \frac{1}{9} \sum_{i=4}^{12} \hat{e}_{t,i}, \quad \text{with} \\ \hat{e}_{t,i} &= y_t - \hat{\beta}_{0,i} - \hat{\beta}_{1,i} y_{t-i} - \hat{\beta}_{2,i} y_{t-i-1} \\ &\quad - \hat{\beta}_{3,i} y_{t-i-2} - \hat{\beta}_{4,i} y_{t-i-3} \end{aligned} \quad (10)$$

Quast and Wolters filter is the third method we use. The fourth method we use is another univariate filter, [Baxter and King \(1999\)](#) filter. In line with [Woitek \(1998\)](#), the impact of this filter on the time series  $y_t$  can be represented as the following moving-average model:

$$z_t = \sum_{j=-K}^K a_j L^j y_t \quad (11)$$

Where  $L$  is the lag operator defined as

$$L^n y_t = y_{t-n} \quad (12)$$

$a_j$  are the weights, and  $z_t$  is the transformed (filtered) time series. The weights  $a_j$  can be calculated by solving the following constrained minimization problem:

$$\min_{a_j} Q = \int_{-\pi}^{\pi} |\beta(\omega) - \alpha(\omega)|^2 d\omega, \text{ s.t. } \alpha(0) = 0 \quad (13)$$

The solution to this minimization problem is:

$$a_j = b_j + \theta; j = 0, \pm 1, \dots, \pm K \quad (14)$$

$$b_j = \begin{cases} \frac{\omega_2 - \omega_1}{\pi} & \text{if } j = 0 \\ \frac{1}{\pi j} (\sin \omega_2 j - \sin \omega_1 j) & \text{if } j = \pm 1, \pm 2, \dots \end{cases} \quad (15)$$

$$\theta = \frac{-\sum_{j=-K}^K b_j}{2K + 1} \quad (16)$$

Parameters  $\omega_1$  and  $\omega_2$  are calculated as:

$$\omega_1 = \frac{2\pi}{p_G}, \omega_2 = \frac{2\pi}{p_D} \quad (17)$$

Where  $p_G$  is the upper bound on the duration of the business cycle, and  $p_D$  is the lower bound. In line with the original [Baxter and King \(1999\)](#) paper, the upper bound is set to 32 quarters, and the lower bound to 6 quarters, while the parameter  $K$  is set to 12, thereby defining the number of forward and backward lags in the equations above. After deciding upon the values of  $p_G$ ,  $p_D$  and  $K$ , one can calculate the values of all variables in equations 6,

7, 8 and 9, which allows for the calculations of weights in equation 3. This transformation of the original time series enables the series to be decomposed into trend and cycle components, that is, real GDP is decomposed into potential GDP and the output gap.

The fifth method we use is the [Christiano and Fitzgerald \(1999\)](#) filter, which belongs to the same filter class as the Baxter-King filter, band-pass filters. The Christiano-Fitzgerald filter is also used to decompose the time series into trend and cycle components, and like the Baxter-King filter, it transforms the original data series using a moving-average model. Before filtering, it is necessary to define the lower and upper limits of the cycle duration, and Christiano and Fitzgerald talk about business cycles as periods of fluctuations lasting between 1, 5 and 8 years. Due to its similarities with the Baxter-King filter and to preserve space, the exact mathematical procedure behind the Christiano-Fitzgerald filtering isn't shown here but is available in the original [Christiano and Fitzgerald \(1999\)](#) paper, as well as publications such as [EU Independent Fiscal Institutions \(2020\)](#).

The sixth method we use is the refined version of the Beveridge-Nelson filter based on the [Kamber et al. \(2018\)](#) and [Kamber et al. \(2025\)](#) papers. [Kamber et al. \(2018\)](#) refine the traditional Beveridge-Nelson filter by imposing a lower signal-to-noise ratio than in the original paper due to the fact that the original Beveridge-Nelson decomposition produces estimates of the output gap that are strongly at odds with widely held beliefs about transitory movements in economic activity. Beveridge-Nelson filter modifications related to Covid recession are further explored in the [Kamber et al. \(2025\)](#).

The seventh method used is the UC-0 model suggested by [Harvey \(1985\)](#) and [Clark \(1987\)](#). The model assumes zero correlation (covariance) between innovations in trend and cycle components. The trend is modeled as a random walk and the cycle as an AR(2) process:

$$\begin{aligned}\tau_t &= \mu_t + \tau_{t-1} + \eta_t \\ c_t &= \varphi_1 c_{t-1} + \varphi_2 c_{t-2} + \varepsilon_t\end{aligned}\tag{18}$$

The first equation represents the trend, denoted by  $\tau_t$ , while the second equation describes the behavior of the cycle as an AR(2) process. The cycle is modeled as an AR(2) process due to the persistence of GDP, and the AR(2) process has proven to be the best choice for developed countries ([Morley et al. \(2003\)](#); [Perron and Wada \(2009\)](#)). Let  $W_t$  be a matrix

containing the innovations of the trend and the cycle,  $\tau_t$  and  $\varepsilon_t$ . Its product yields the variance–covariance matrix:

$$E(W_t W_t') = \begin{bmatrix} \sigma_\eta^2 & \sigma_{\eta\varepsilon} \\ \sigma_{\varepsilon\eta} & \sigma_\varepsilon^2 \end{bmatrix} \quad (19)$$

$W_t W_t'$  is a symmetric matrix, implying  $\sigma_{\eta\varepsilon} = \sigma_{\varepsilon\eta}$ . Standard models assume that  $\sigma_{\eta\varepsilon} = 0$ , which corresponds to a model with restrictions on the covariance between the trend and the cycle.

The eight, ninth and tenth output gap estimation methods are based on a bivariate SVAR model. The first SVAR model uses employment, the second unemployment and the third employs inflation together with GDP in the vector of endogenous variables:  $\Gamma_t = [\Delta y_t, n_t]$ ,  $\Gamma_t = [\Delta y_t, u_t]$  or  $\Gamma_t = [\Delta y_t, \pi_t]$ :

$$\Gamma_t = A_0 + A_1 \Gamma_{t-1} + \dots + A_p \Gamma_{t-p} + \mathcal{E}_t, \quad (20)$$

where  $\Delta y_t$  is the GDP growth rate,  $u_t$  is the unemployment to labor force ratio,  $n_t$  is the ratio of employment to working age population,  $\pi_t$  is the inflation rate,  $A_0$  is the vector of constant terms, and  $\mathcal{E}_t$  is the vector of error terms.  $A_1, \dots, A_p$  are matrices of coefficients for  $p$  lags of  $\Gamma$ , and we uniformly estimate  $p = 4$  for all countries in our sample. Long-run restrictions à la [Blanchard and Quah \(1989\)](#) are imposed to identify supply and demand shocks, where demand shocks do not affect GDP growth in the long run. Potential output is calculated as GDP in the absence of demand shocks using a historical decomposition of the GDP growth rate to decompose its movements into supply and demand shocks. The output gap is thus the difference between the actual and potential GDP, or, alternatively, the overall demand shocks estimated using the historical decomposition. The inclusion of the inflation rate as the second variable (the last SVAR model) follows the work of [Bayoumi and Eichengreen \(1992\)](#), which can be considered an extension of the original [Blanchard and Quah \(1989\)](#) paper.

Figure A.1: Gross Domestic Product

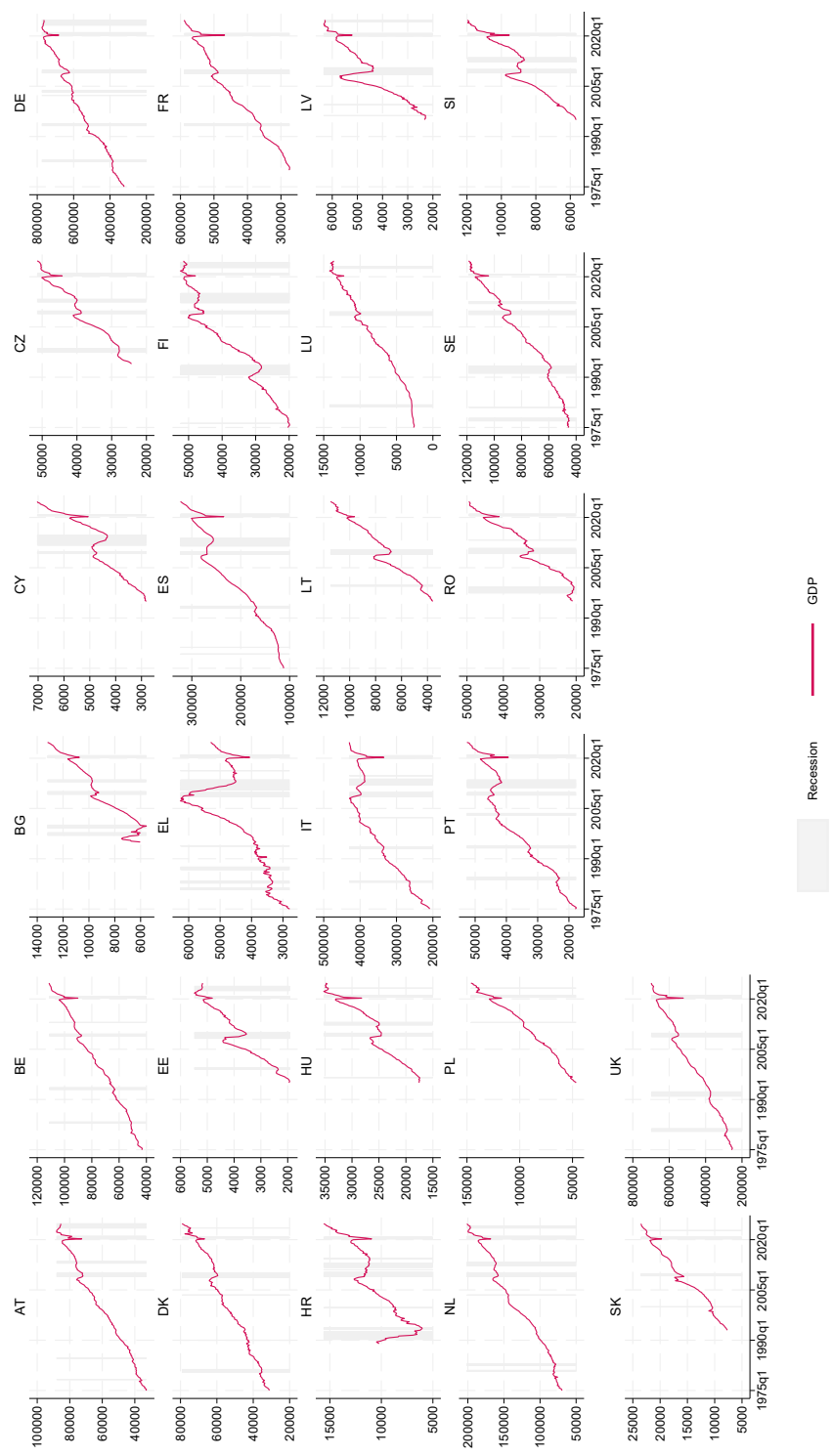


Figure A.2: Gross domestic product growth rate and GDP deflator inflation rate

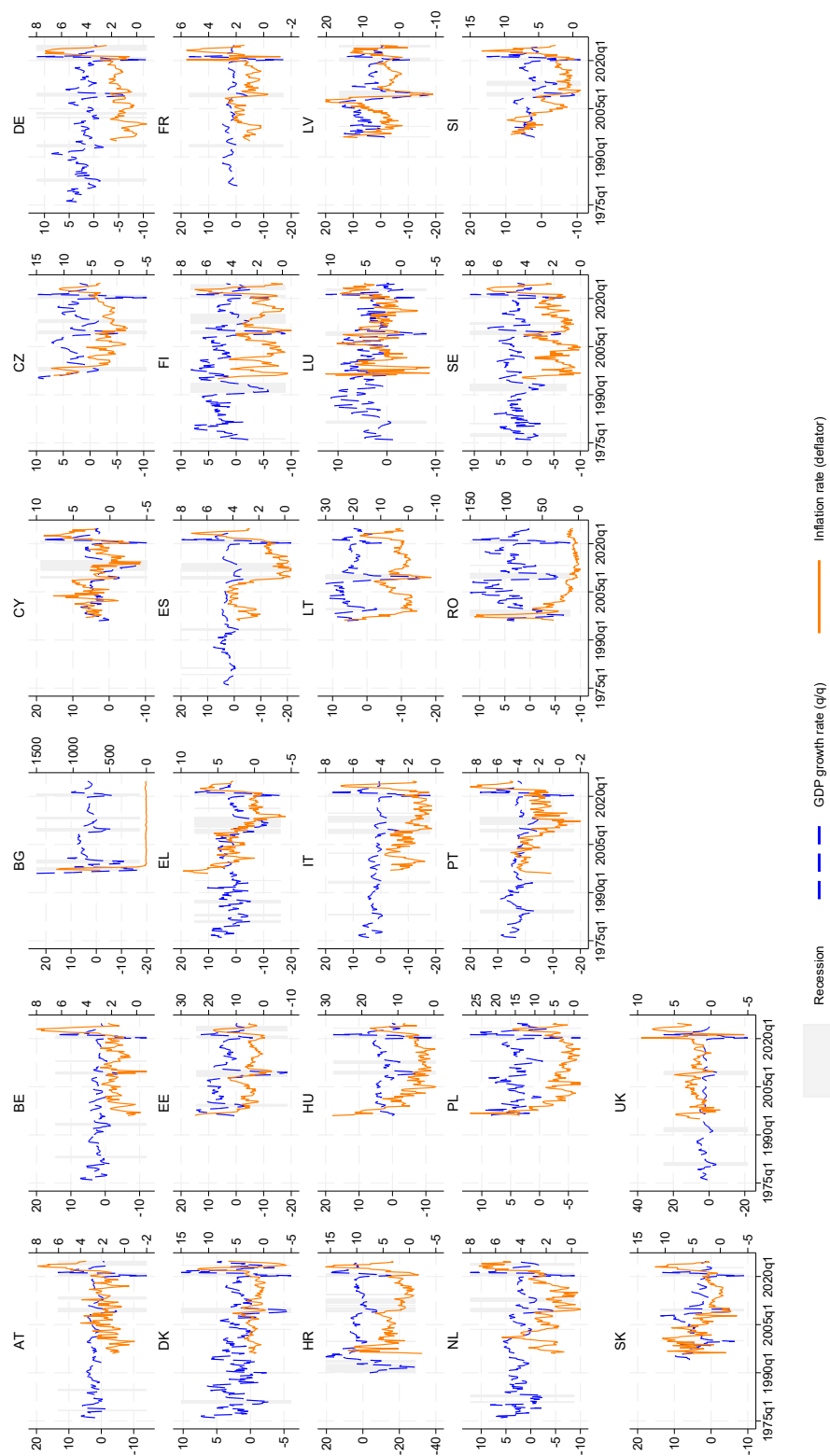
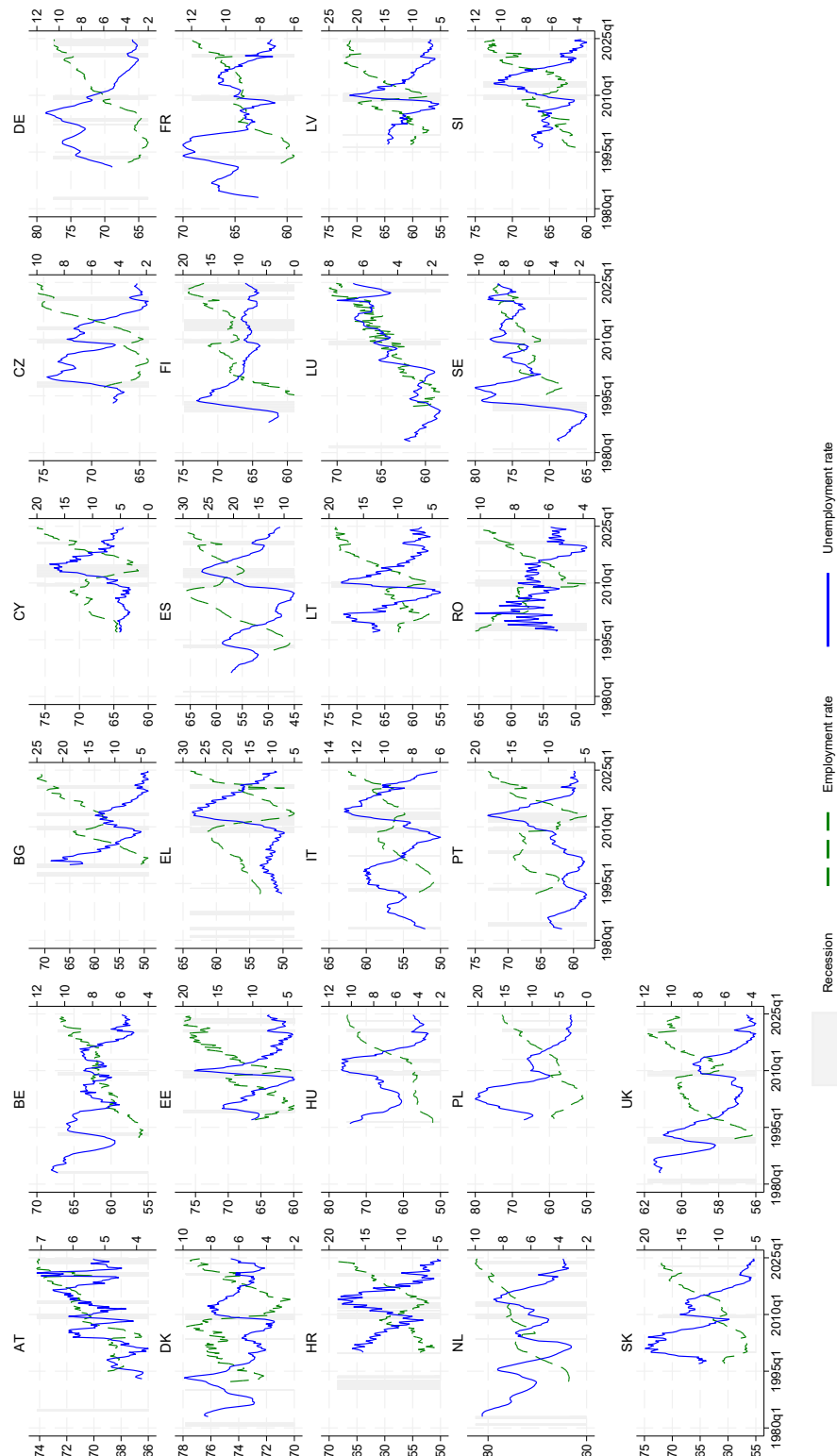


Figure A.3: Employment and unemployment rate



## Appendix B: Robustness check

Table B.1: MAOG as the threshold variable, nonlinear model with quasi-real-time output gap estimates

|                 | (1)<br>ReSD, d=0     | (2)<br>ReSD, d=0     | (3)<br>ReSD, d=0     | (4)<br>ReSD, d=1     | (5)<br>ReSD, d=1     | (6)<br>ReSD, d=1     |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy_1            | -0.234***<br>(0.000) | -0.212***<br>(0.000) | -0.118***<br>(0.001) | -0.123***<br>(0.000) | -0.115***<br>(0.000) | -0.070***<br>(0.000) |
| lag4dependent_1 |                      | 0.541***<br>(0.000)  | 0.190**<br>(0.028)   |                      | 0.528***<br>(0.000)  | 0.135**<br>(0.038)   |
| lag1dependent_1 |                      |                      | 0.613***<br>(0.000)  |                      |                      | 0.693***<br>(0.000)  |
| dy_2            | 0.002<br>(0.964)     | -0.016<br>(0.367)    | -0.000<br>(0.963)    | 0.020<br>(0.732)     | -0.039**<br>(0.044)  | -0.010<br>(0.127)    |
| lag4dependent_2 |                      | 0.560***<br>(0.000)  | -0.065**<br>(0.022)  |                      | 0.629***<br>(0.000)  | -0.088**<br>(0.012)  |
| lag1dependent_2 |                      |                      | 0.915***<br>(0.000)  |                      |                      | 0.939***<br>(0.000)  |
| _cons           | 1.945***<br>(0.000)  | 0.527***<br>(0.008)  | -0.084<br>(0.576)    | 1.895***<br>(0.000)  | 0.874***<br>(0.000)  | 0.000<br>(0.999)     |
| <i>N</i>        | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| Tr_value        | -2.491               | -2.491               | -2.153               | 1.630                | 1.630                | -0.406               |
| Wald_pval       | 0.000                | 0.000                | 0.002                | 0.013                | 0.001                | 0.003                |
| rmse            | 0.880                | 0.710                | 0.481                | 0.894                | 0.723                | 0.490                |
| aic             | 6130.445             | 4882.941             | 3099.576             | 6206.089             | 4965.105             | 3182.036             |
| bic             | 6274.942             | 5026.327             | 3242.962             | 6350.586             | 5108.491             | 3325.422             |

**Notes:** *p*-values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ . In regime 1, below threshold value  $\theta$ ,  $dy_1$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_1 is  $\delta_{it-1}$  and lag4dependent\_1 is  $\delta_{it-4}$ . In regime 2, above threshold value  $\theta$ ,  $dy_2$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_2 is  $\delta_{it-1}$  and lag4dependent\_2 is  $\delta_{it-4}$ . *Tr\_value* presents estimated threshold values  $\theta$ . *Wald\_pval* is the *p*-value for the Wald test with  $H_0 : dy_1 = dy_2$ . Threshold variable  $q_{it-d}$  is  $c_{it-d}$  and delay parameter  $d$  is lag of the threshold variable. Threshold grid search results are presented in Figure B.2.

Table B.2: Absolute value of the GDP growth rate as the threshold variable, nonlinear model with quasi-real-time output gap estimates

|                  | (1)<br>ReSD, d=0     | (2)<br>ReSD, d=0     | (3)<br>ReSD, d=0     | (4)<br>ReSD, d=1     | (5)<br>ReSD, d=1     | (6)<br>ReSD, d=1     |
|------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy_1             | -0.155***<br>(0.000) | -0.066***<br>(0.004) | -0.035***<br>(0.001) | -0.132***<br>(0.000) | -0.094***<br>(0.000) | -0.058***<br>(0.000) |
| lag4dependent_1  |                      | 0.460***<br>(0.000)  | 0.005<br>(0.856)     |                      | 0.502***<br>(0.000)  | -0.009<br>(0.637)    |
| lag1dependent_1  |                      |                      | 0.797***<br>(0.000)  |                      |                      | 0.893***<br>(0.000)  |
| dy_2             | -0.086***<br>(0.006) | -0.104***<br>(0.000) | -0.040***<br>(0.001) | -0.080**<br>(0.018)  | -0.093***<br>(0.000) | -0.020**<br>(0.037)  |
| lag4dependent_2  |                      | 0.673***<br>(0.000)  | 0.108<br>(0.133)     |                      | 0.631***<br>(0.000)  | 0.119<br>(0.141)     |
| lag1dependent_2  |                      |                      | 0.758***<br>(0.000)  |                      |                      | 0.725***<br>(0.000)  |
| _cons            | 2.241***<br>(0.000)  | 0.776***<br>(0.001)  | 0.070<br>(0.670)     | 2.215***<br>(0.000)  | 0.758***<br>(0.000)  | 0.008<br>(0.959)     |
| <i>N</i>         | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| <i>Tr_value</i>  | 3.196                | 4.842                | 4.842                | 3.439                | 4.861                | 6.335                |
| <i>Wald_pval</i> | 0.055                | 0.126                | 0.695                | 0.117                | 0.947                | 0.001                |
| <i>rmse</i>      | 0.923                | 0.722                | 0.500                | 0.922                | 0.738                | 0.492                |
| <i>aic</i>       | 6357.991             | 4958.090             | 3277.487             | 6354.063             | 5056.292             | 3205.690             |
| <i>bic</i>       | 6502.488             | 5101.476             | 3420.873             | 6498.560             | 5199.678             | 3349.075             |

**Notes:** *p*-values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ . In regime 1, below threshold value  $\theta$ ,  $dy_1$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_1 is  $\delta_{it-1}$  and lag4dependent\_1 is  $\delta_{it-4}$ . In regime 2, above threshold value  $\theta$ ,  $dy_2$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_2 is  $\delta_{it-1}$  and lag4dependent\_2 is  $\delta_{it-4}$ . *Tr\_value* presents estimated threshold values  $\theta$ . *Wald\_pval* is the *p*-value for the Wald test with  $H_0 : dy_1 = dy_2$ . Threshold variable  $q_{it-d}$  is  $|\Delta y_{it-d}|$  and delay parameter *d* is lag of the threshold variable. Threshold grid search results are presented in Figure B.3.

Table B.3: GDP growth rate as the threshold variable, nonlinear model with quasi-real-time output gap estimates (Uncertainty is calculated without the IMF, OECD and EC output gaps)

|                 | (1)<br>ReSD, d=0     | (2)<br>ReSD, d=0     | (3)<br>ReSD, d=0     | (4)<br>ReSD, d=1     | (5)<br>ReSD, d=1     | (6)<br>ReSD, d=1     |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy_1            | -0.345***<br>(0.000) | -0.365***<br>(0.000) | -0.274***<br>(0.000) | -0.213***<br>(0.000) | -0.232***<br>(0.000) | -0.168***<br>(0.000) |
| lag4dependent_1 |                      | 0.357***<br>(0.000)  | 0.169<br>(0.108)     |                      | 0.519***<br>(0.000)  | 0.248***<br>(0.010)  |
| lag1dependent_1 |                      |                      | 0.427***<br>(0.000)  |                      |                      | 0.488***<br>(0.000)  |
| dy_2            | 0.055<br>(0.291)     | -0.007<br>(0.824)    | -0.004<br>(0.758)    | -0.016<br>(0.661)    | -0.028<br>(0.274)    | -0.013<br>(0.179)    |
| lag4dependent_2 |                      | 0.541***<br>(0.000)  | -0.032*<br>(0.076)   |                      | 0.547***<br>(0.000)  | -0.074***<br>(0.003) |
| lag1dependent_2 |                      |                      | 0.885***<br>(0.000)  |                      |                      | 0.931***<br>(0.000)  |
| _cons           | 1.791***<br>(0.000)  | 0.549**<br>(0.014)   | -0.106<br>(0.549)    | 2.004***<br>(0.000)  | 0.599***<br>(0.004)  | -0.085<br>(0.609)    |
| <i>N</i>        | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| Tr_value        | 0.029                | -0.131               | -0.993               | -1.015               | -1.015               | -1.015               |
| Wald_pval       | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                | 0.000                |
| rmse            | 0.935                | 0.774                | 0.538                | 0.975                | 0.801                | 0.551                |
| aic             | 6422.479             | 5274.465             | 3613.695             | 6622.173             | 5435.773             | 3720.511             |
| bic             | 6566.977             | 5417.850             | 3757.080             | 6766.670             | 5579.159             | 3863.897             |

**Notes:** *p*-values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ . In regime 1, below threshold value  $\theta$ ,  $dy_1$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_1 is  $\delta_{it-1}$  and lag4dependent\_1 is  $\delta_{it-4}$ . In regime 2, above threshold value  $\theta$ ,  $dy_2$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_2 is  $\delta_{it-1}$  and lag4dependent\_2 is  $\delta_{it-4}$ . *Tr\_value* presents estimated threshold values  $\theta$ . *Wald\_pval* is the *p*-value for the Wald test with  $H_0 : dy_1 = dy_2$ . Threshold variable  $q_{it-d}$  is  $\Delta y_{it-d}$  and delay parameter *d* is lag of the threshold variable. Threshold grid search results are presented in Figure B.4.

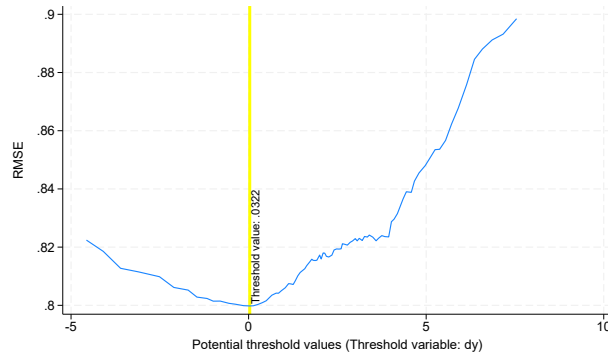
Table B.4: GDP growth rate as the threshold variable, nonlinear model with quasi-real-time output gap estimates (Uncertainty is calculated without the UC and Hamilton output gaps)

|                 | (1)<br>ReSD, d=0     | (2)<br>ReSD, d=0     | (3)<br>ReSD, d=0     | (4)<br>ReSD, d=1     | (5)<br>ReSD, d=1     | (6)<br>ReSD, d=1     |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy_1            | -0.272***<br>(0.000) | -0.243***<br>(0.000) | -0.159***<br>(0.000) | -0.157***<br>(0.000) | -0.157***<br>(0.000) | -0.112***<br>(0.000) |
| lag4dependent_1 |                      | 0.473***<br>(0.000)  | 0.217**<br>(0.044)   |                      | 0.569***<br>(0.000)  | 0.238***<br>(0.001)  |
| lag1dependent_1 |                      |                      | 0.519***<br>(0.000)  |                      |                      | 0.540***<br>(0.000)  |
| dy_2            | 0.023<br>(0.614)     | -0.006<br>(0.814)    | 0.002<br>(0.808)     | -0.031<br>(0.445)    | -0.026<br>(0.179)    | -0.004<br>(0.465)    |
| lag4dependent_2 |                      | 0.560***<br>(0.000)  | -0.061***<br>(0.000) |                      | 0.574***<br>(0.000)  | -0.103***<br>(0.000) |
| lag1dependent_2 |                      |                      | 0.933***<br>(0.000)  |                      |                      | 0.980***<br>(0.000)  |
| _cons           | 1.884***<br>(0.000)  | 0.156<br>(0.425)     | -0.449***<br>(0.009) | 2.177***<br>(0.000)  | 0.201<br>(0.278)     | -0.437**<br>(0.012)  |
| <i>N</i>        | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| Tr_value        | -0.131               | -0.131               | -0.993               | 3.646                | 0.281                | -1.015               |
| Wald_pval       | 0.000                | 0.000                | 0.000                | 0.002                | 0.000                | 0.000                |
| rmse            | 0.819                | 0.657                | 0.436                | 0.844                | 0.671                | 0.439                |
| aic             | 5784.157             | 4530.078             | 2645.165             | 5933.355             | 4623.045             | 2680.322             |
| bic             | 5928.654             | 4673.464             | 2788.551             | 6077.852             | 4766.431             | 2823.708             |

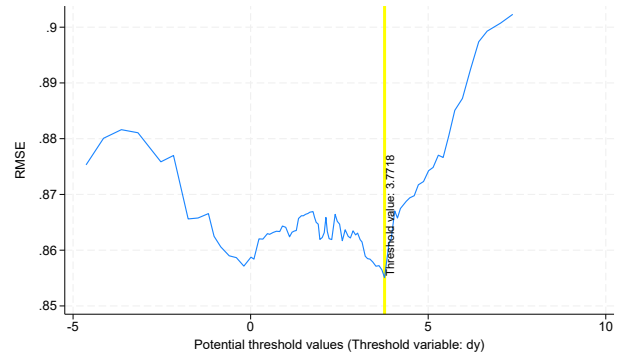
**Notes:**  $p$ -values for estimated coefficients are in parentheses. \*, \*\* and \*\*\* indicate statistical significance at 10%, 5% and 1% levels respectively. Dependent variable is uncertainty  $\delta_{it}$ . In regime 1, below threshold value  $\theta$ ,  $dy_1$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_1 is  $\delta_{it-1}$  and lag4dependent\_1 is  $\delta_{it-4}$ . In regime 2, above threshold value  $\theta$ ,  $dy_2$  is GDP growth rate  $\Delta y_{it}$ , lag1dependent\_2 is  $\delta_{it-1}$  and lag4dependent\_2 is  $\delta_{it-4}$ . *Tr\_value* presents estimated threshold values  $\theta$ . *Wald\_pval* is the  $p$ -value for the Wald test with  $H_0 : dy_1 = dy_2$ . Threshold variable  $q_{it-d}$  is  $\Delta y_{it-d}$  and delay parameter  $d$  is lag of the threshold variable. Threshold grid search results are presented in Figure B.5.

Figure B.1: Estimated RMSE vs potential threshold values (Grid search for threshold values performed between 5th and 95th percentile of threshold variable, GDP growth rate as threshold variable)

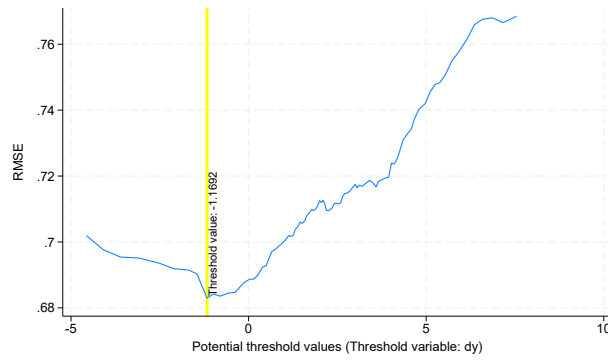
(a) *Model 1: RMSE vs threshold values (delay=0)*



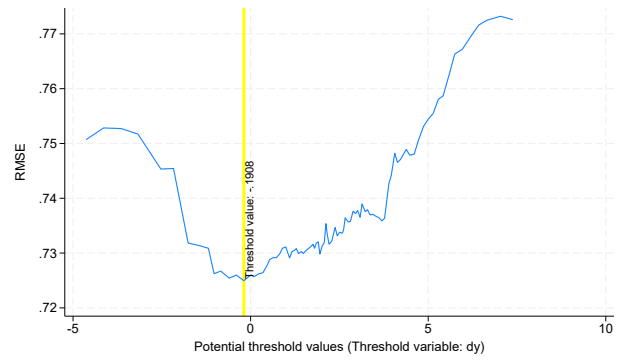
(b) *Model 4: RMSE vs threshold values (delay=1)*



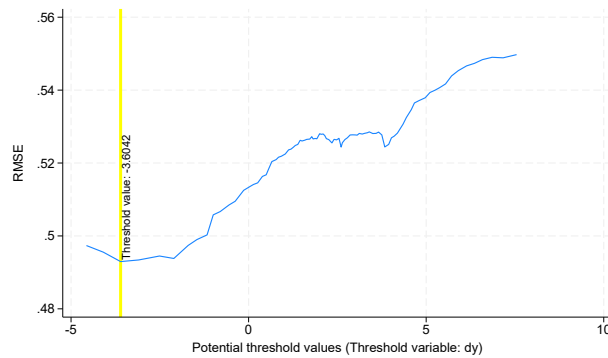
(c) *Model 2: RMSE vs threshold values (delay=0)*



(d) *Model 5: RMSE vs threshold values (delay=1)*



(e) *Model 3: RMSE vs threshold values (delay=0)*



(f) *Model 6: RMSE vs threshold values (delay=1)*

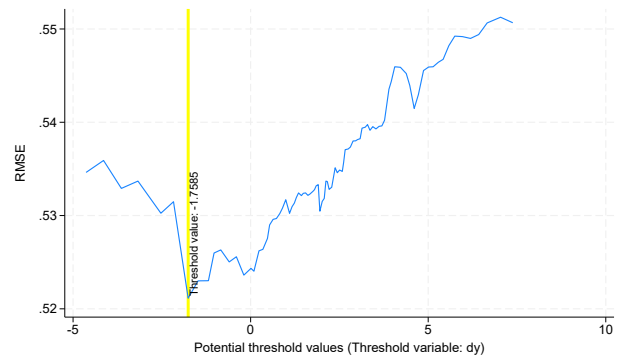
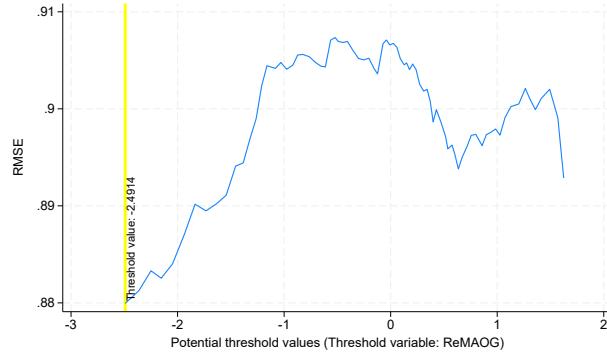
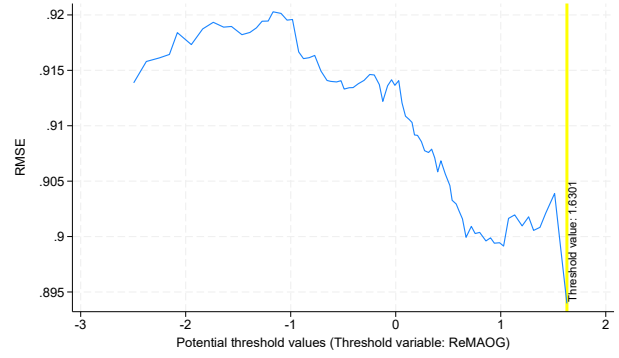


Figure B.2: Estimated RMSE vs potential threshold values (MAOG as threshold variable)

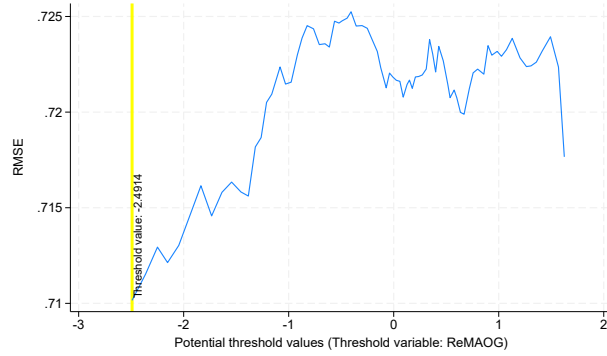
(a) *Model 1: RMSE vs threshold values (delay=0)*



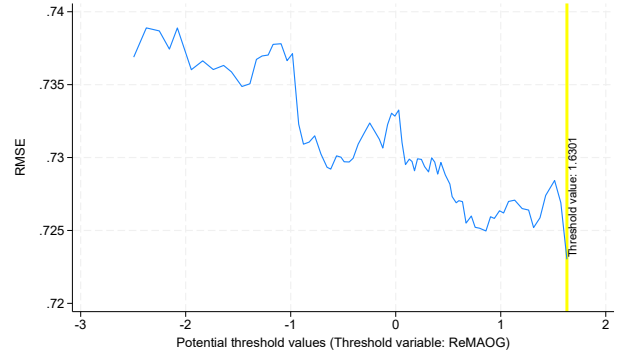
(b) *Model 4: RMSE vs threshold values (delay=1)*



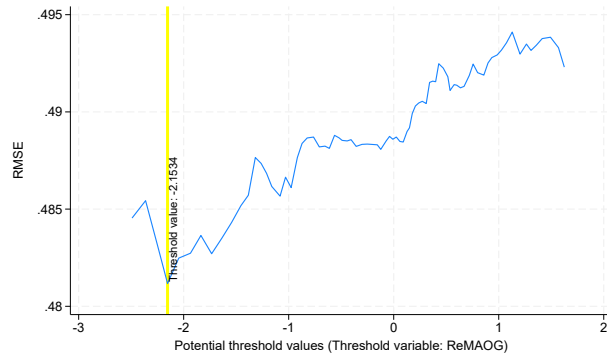
(c) *Model 2: RMSE vs threshold values (delay=0)*



(d) *Model 5: RMSE vs threshold values (delay=1)*



(e) *Model 3: RMSE vs threshold values (delay=0)*



(f) *Model 6: RMSE vs threshold values (delay=1)*

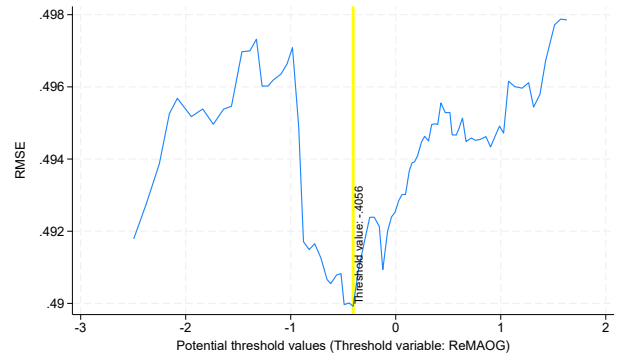
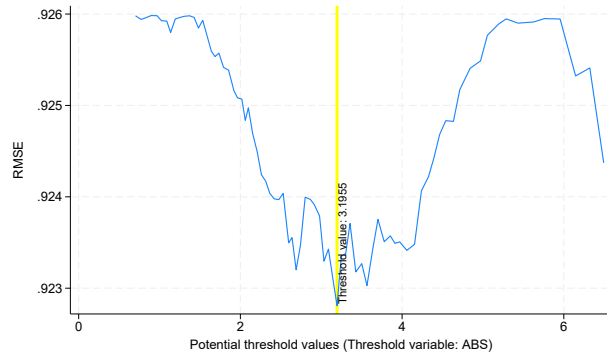
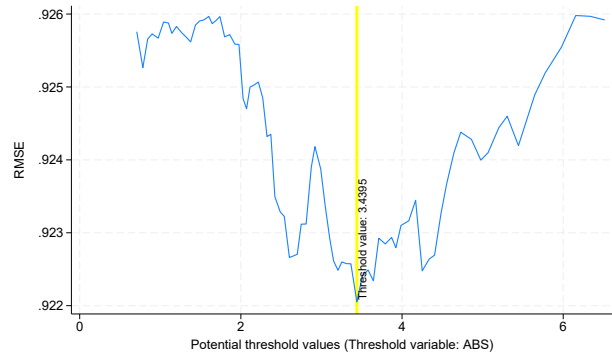


Figure B.3: Estimated RMSE vs potential threshold values (Absolute value of GDP growth as threshold variable)

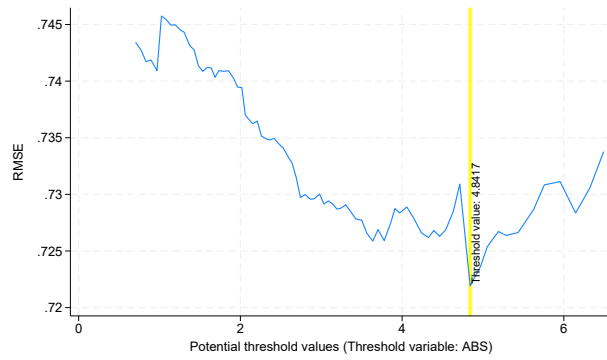
(a) Model 1: RMSE vs threshold values (delay=0)



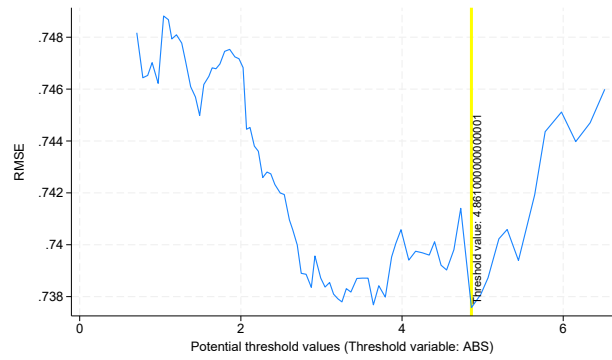
(b) Model 4: RMSE vs threshold values (delay=1)



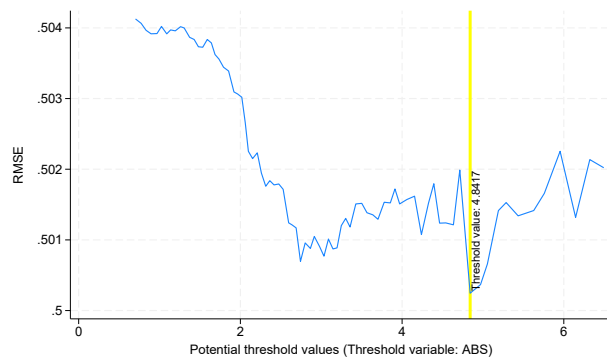
(c) Model 2: RMSE vs threshold values (delay=0)



(d) Model 5: RMSE vs threshold values (delay=1)



(e) Model 3: RMSE vs threshold values (delay=0)



(f) Model 6: RMSE vs threshold values (delay=1)

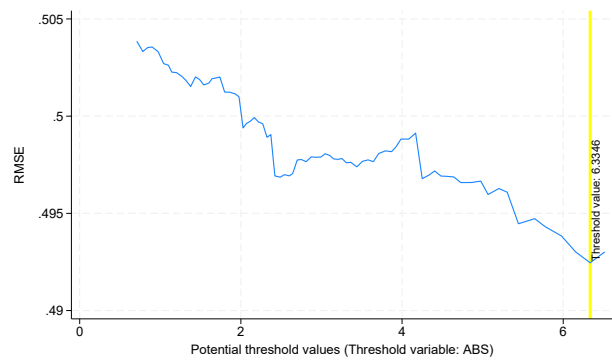
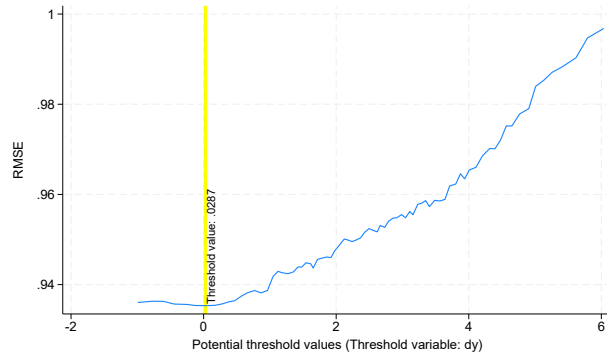
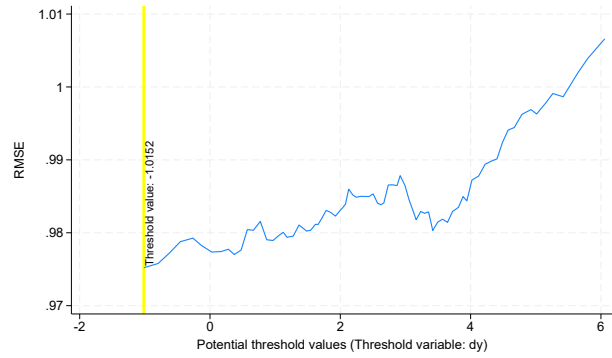


Figure B.4: Estimated RMSE vs potential threshold values (Uncertainty is calculated without IMF, OECD and EC output gap, GDP growth rate as threshold variable)

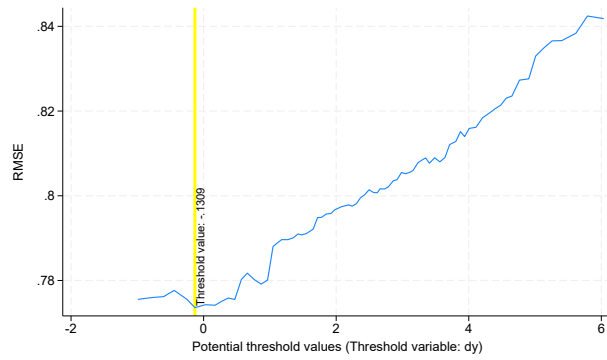
(a) *Model 1: RMSE vs threshold values (delay=0)*



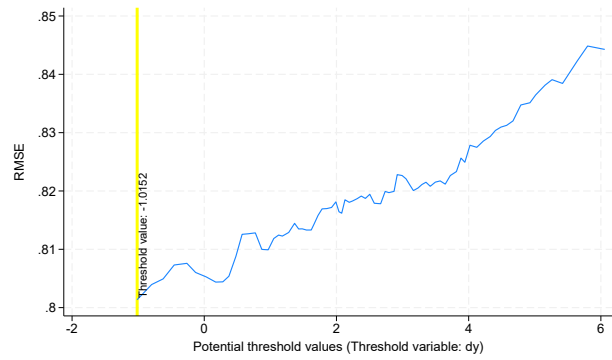
(b) *Model 4: RMSE vs threshold values (delay=1)*



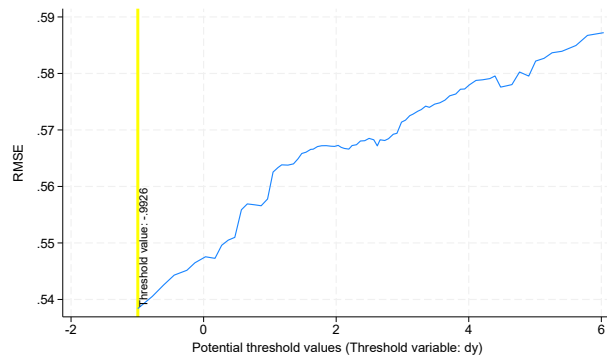
(c) *Model 2: RMSE vs threshold values (delay=0)*



(d) *Model 5: RMSE vs threshold values (delay=1)*



(e) *Model 3: RMSE vs threshold values (delay=0)*



(f) *Model 6: RMSE vs threshold values (delay=1)*

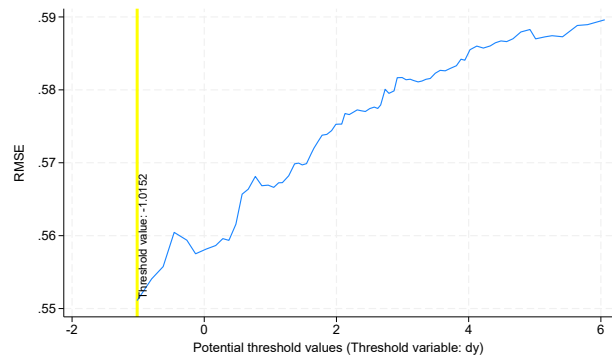
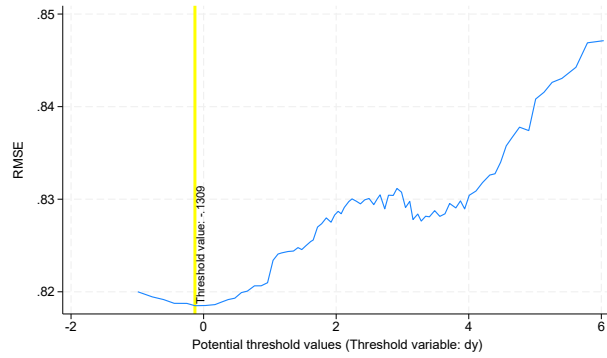
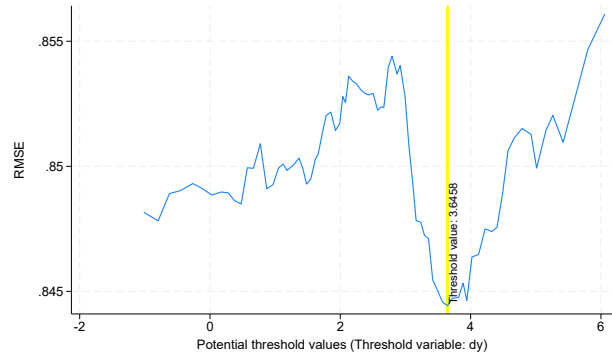


Figure B.5: Estimated RMSE vs potential threshold values (Uncertainty is calculated without UC and Hamilton output gap, GDP growth rate as threshold variable)

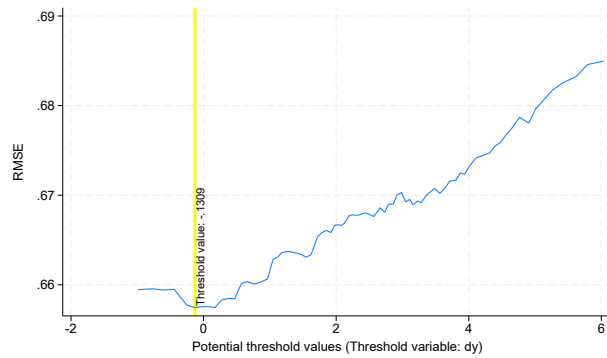
(a) *Model 1: RMSE vs threshold values (delay=0)*



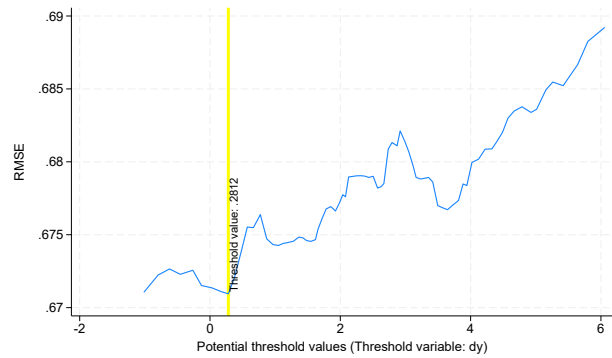
(b) *Model 4: RMSE vs threshold values (delay=1)*



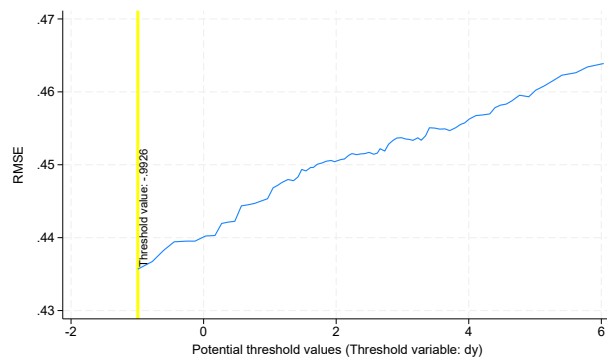
(c) *Model 2: RMSE vs threshold values (delay=0)*



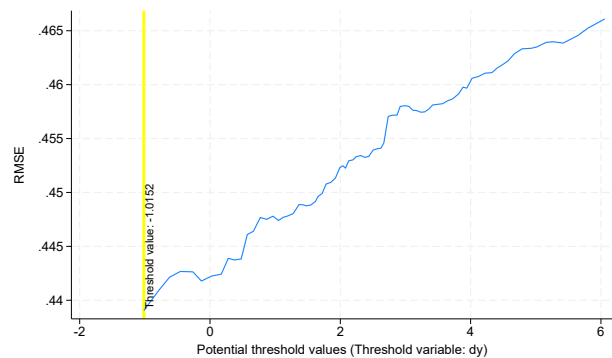
(d) *Model 5: RMSE vs threshold values (delay=1)*



(e) *Model 3: RMSE vs threshold values (delay=0)*



(f) *Model 6: RMSE vs threshold values (delay=1)*



# Appendix C: Online appendix with results of final output gaps (not to be published)

Table C.1: Linear model with final output gap estimates

|               | (1)<br>FeSD          | (2)<br>FeSD          | (3)<br>FeSD          | (4)<br>FeSD          | (5)<br>FeSD          | (6)<br>FeSD          | (7)<br>FeSD          | (8)<br>FeSD          | (9)<br>FeSD          |
|---------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| dy            | -0.044***<br>(0.000) | -0.070***<br>(0.000) | -0.032***<br>(0.000) | -0.065***<br>(0.000) | -0.073***<br>(0.000) | -0.036***<br>(0.000) | -0.057***<br>(0.001) | -0.065***<br>(0.000) | -0.024***<br>(0.001) |
| lag1dependent |                      |                      | 0.761***<br>(0.000)  |                      |                      | 0.708***<br>(0.000)  |                      |                      | 0.861***<br>(0.000)  |
| lag4dependent |                      | 0.858***<br>(0.000)  | 0.186***<br>(0.000)  |                      | 0.669***<br>(0.000)  | 0.152***<br>(0.002)  |                      | 0.699***<br>(0.000)  | 0.038<br>(0.247)     |
| _cons         | 2.851***<br>(0.000)  | 0.516***<br>(0.000)  | 0.204***<br>(0.000)  | 2.896***<br>(0.000)  | 1.052***<br>(0.000)  | 0.455***<br>(0.000)  | 2.768***<br>(0.000)  | 1.000***<br>(0.000)  | 0.357***<br>(0.000)  |
| <i>N</i>      | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 | 2392                 | 2288                 | 2288                 |
| countryFE     |                      |                      |                      | Yes                  | Yes                  | Yes                  | Yes                  | Yes                  | Yes                  |
| timeFE        |                      |                      |                      |                      |                      |                      | Yes                  | Yes                  | Yes                  |
| rmse          | 1.408                | 0.715                | 0.502                | 0.923                | 0.665                | 0.490                | 0.795                | 0.545                | 0.339                |
| aic           | 8428.134             | 4959.349             | 3343.443             | 6405.958             | 4627.941             | 3232.358             | 5644.601             | 3672.607             | 1494.615             |
| bic           | 8439.694             | 4976.555             | 3366.384             | 6411.738             | 4639.411             | 3249.565             | 5789.098             | 3815.993             | 1638.001             |

Figure C.1: Final output gaps

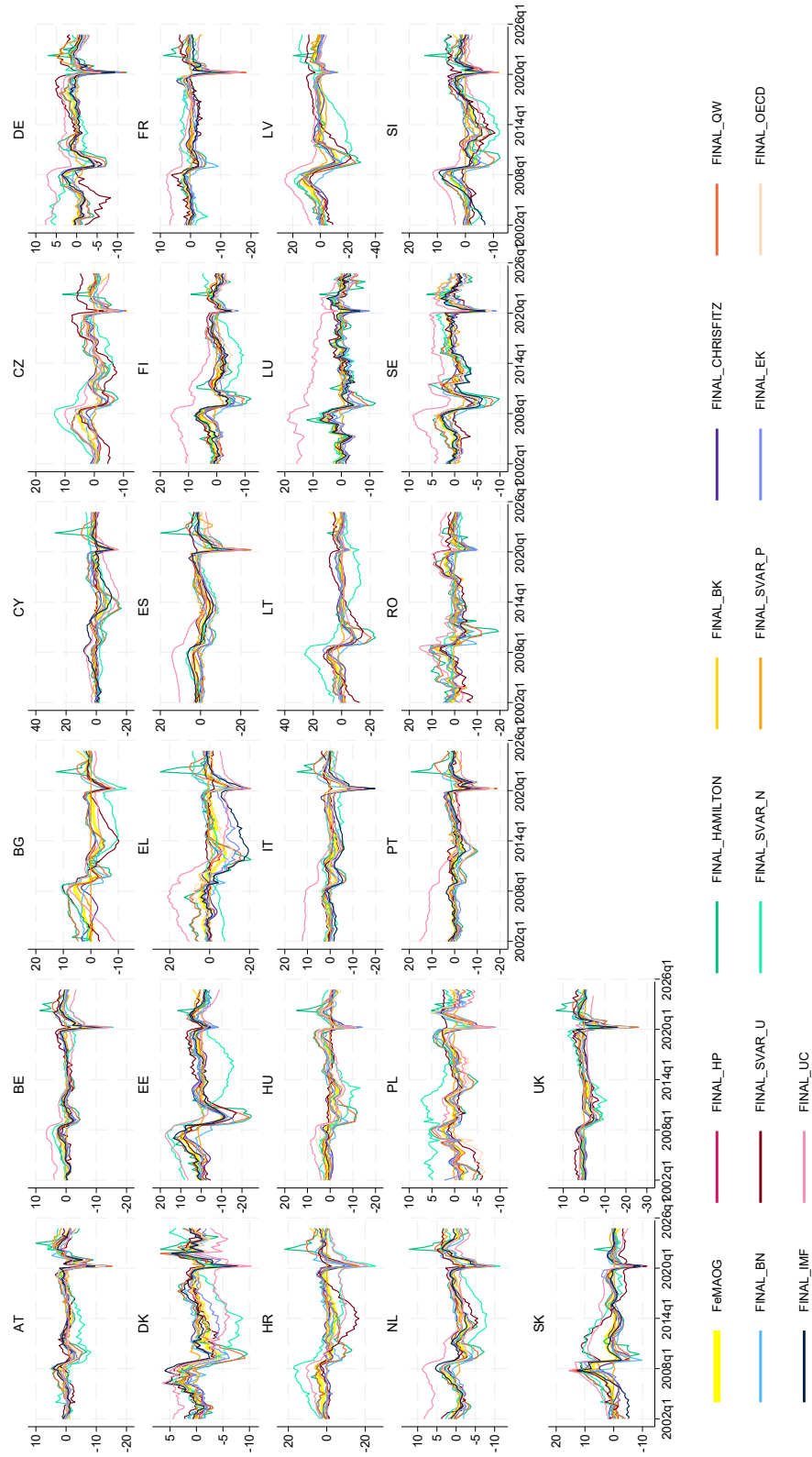


Figure C.2: Standard deviation for the quasi-real-time and final output gap estimates

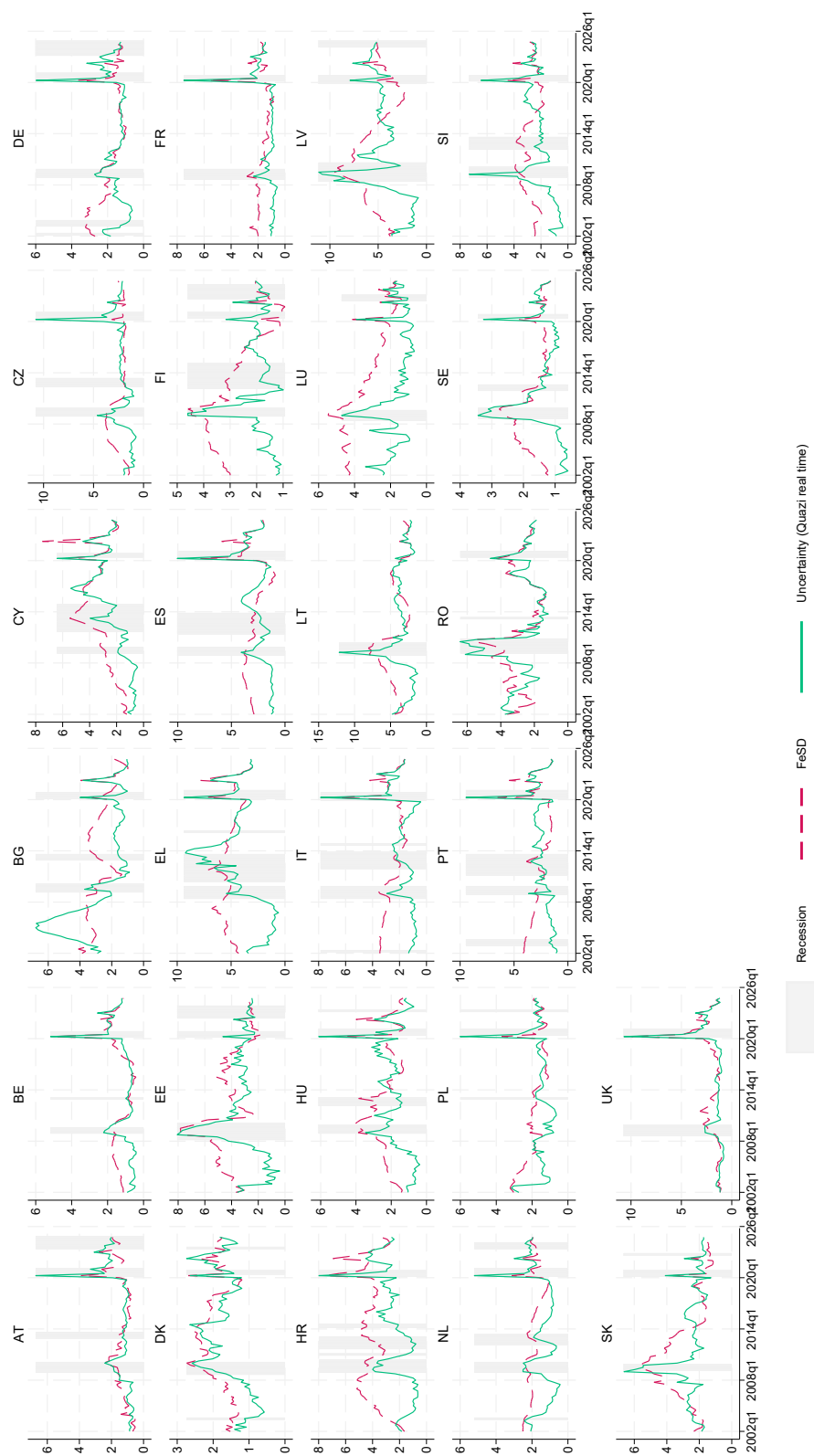


Figure C.3: Standard deviation decomposition by output gap methods – Final gaps

