EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Máté Tóth

A multivariate unobserved components model to estimate potential output in the euro area: a production function based approach

No 2523 / February 2021

Abstract

This paper builds an unobserved components model that combines a multivariate filter approach with a Cobb-Douglas production function. This combination allows potential output estimates to incorporate more economic structure than the traditional production function approach, while retaining the ability to conduct growth accounting exercises. The model is a backward-looking state space model estimated with Bayesian methods employing the Kalman filter to jointly decompose six key observable variables (real GDP, unemployment rate, labour force participation rate, hours worked per person, a measure of core inflation and wage inflation) into trend and cyclical components. To do so, it relies on several reduced form relationships across the cyclical components, such as a wage and a price Phillips curve and an Okun's law type relationship, while it also assumes common trends for a few variables and allows for hysteresis effects. The model is estimated on aggregate euro area data with Bayesian methods. The paper finds that the resulting output gap estimates have good revision properties and reasonable forecasting performance in particular in terms of GDP and core inflation *vis-a-vis* a set of benchmarks.

JEL classification: C32, D24, E32, E37

Keywords: state-space model, production function, Bayesian estimation

Non-technical summary

The conduct of monetary policy relies on the assessment of the state of the economy which is often defined as the deviation of activity from its long-run trend, or the business cycle. However, the long-run trend of economic activity and thus the business cycle is unobservable, i.e. it cannot be directly measured, but has to be inferred from the available data. Unfortunately, there are infinitely many ways to decompose observable data into unobservable trend and cyclical components. The researcher thus needs to find a way to 'discipline' the decomposition by making assumptions about the processes governing trends and cycles. For example, a typical assumption is that the time series of the trend evolves in a smoother way than the cycle. In a multivariate context, assumptions can be made on the basis of economic theory, such as imposing a relationship between price and/or wage pressures and the business cycle.

This paper proposes a method for estimating the unobservable states of the economy that are relevant for monetary policy, such as the output gap (i.e. the difference between actual output and its potential level) or the unemployment gap (the difference between the actual unemployment rate and its trend). In order to avoid the pitfalls of some widely-used trend-cycle decomposition methods, the current approach uses a multi-variate set-up, according to which the extraction of the unobservable states is guided by a set of economic relationships assumed to drive the development of the sates over time, i.e. a model. The model is built around a production function but allows for more economic structure than more traditional approaches while retaining the ability to conduct growth accounting exercises. The model decomposes six key observable variables (real GDP, unemployment rate, labour force participation rate, hours worked per person, core inflation and wage inflation) simultaneously into their trend and cyclical components. To do so, it relies on well-established relationships, such as a wage and a price Phillips curve, connecting wage developments with labour market slack and price developments with the output gap, as well as an Okun's law type relationship connecting labour market slack to the output gap. Importantly, the model also features connected trends and allows for hysteresis effects, i.e. the possibility of transitory fluctuations exerting an effect on long-term trends. The model is estimated on aggregate euro area data via Bayesian methods, a combination of prior beliefs about the prevailing economic relationships and the information in the data.

Such estimates of the state of the economy are always surrounded by a high degree of uncertainty and thus tend to be revised when new information becomes available, e.g the publication of an additional quarter of national accounts data. Given their unobservability, it is not straightforward to evaluate or choose between alternative estimates of the state of the economy. Typically, the estimates are expected to provide a plausible narrative about the evolution of the state of the economy, revised to a limited extent over time in response to incoming data and have some predictive power with respect to the observable variables from which they were derived. The model based estimates of the unobserved state of the economy introduced in this paper are evaluated against the above criteria and are found to have good revision properties and reasonable forecasting performance particularly in terms of GDP growth and core inflation *vis-a-vis* a set of benchmarks.

1 Introduction

Observable time series can be decomposed to unobservable trend and cyclical components in infinitely many ways. In order to achieve a specific decomposition one needs to impose restrictions on the behaviour of the unobserved components of the observable series. These restrictions can reflect some desirable statistical properties of the components, such as the smoothness or the periodicity of the trend component relative to the cycle. In a multivariate setup the assumptions about interrelationships between the unobserved components of interest can serve as additional restrictions. For example, the output gap, the difference between actual and potential output, is typically considered as a key determinant of inflationary pressures in an economy, thus observable inflation may contain useful information about the size of the output gap. Similarly, the unemployment rate is often thought of as a reliable business cycle indicator, see e.g. Ball et al. (2017), and can thus be leveraged to inform output gap estimates. The most widely used methods for estimating the level of key unobservables of macroeconomic relevance, such as potential output, are univariate filters, multivariate filters, semi-structural models such as Benes et al. (2010) and fully-fledged structural models[1].

A common feature of the above methods is that they have (at least an approximate) state-space representation where the Kalman-filter or smoother can be applied to extract the unobservable trend and cyclical components from the observable data. Furthermore, if the unobservable state(s) are estimated with the help of a two-sided filter[2] or smoother, as is the case in typical applications, all the listed methods will suffer from the so-called end point problem: new observations added to the sample will have an impact on the entire path of smoothed state estimates. The severity of revisions due to new observations added to the sample depend on how accurately the model underlying the smoothing problem predicts the data releases, as shown e.g. by Kaiser and Maravall (1999). This follows from the mechanics of the Kalman smoothing algorithm,

[1]New Keynesian DSGE models have been used to calculate the natural level of output, which is a counterfactual concept that is similar to- but not identical to the concept of potential output. The natural level of output refers to an output path under the assumption of no nominal rigidities (such as sticky prices or wages), see e.g. Vetlov et al. (2011). As shown by Primiceri and Justiniano (2009) in an estimated new Keynesian DSGE model, unless mark-up shocks are treated as measurement noise instead of underlying drivers of inflation dynamics, natural output tends to be significantly more volatile than actual output. By assuming away exogenous fluctuation in the monopolistic power of firms and workers (i.e. mark-up shocks) as underlying drivers of wage and price developments, an otherwise standard new Keynesian model yields a relatively smooth path of natural output, that is in line with more traditional definitions of potential output.

[2]A one sided filter in a sample size of τ provides estimate of the unobserved state at time $t \leq \tau$ based on sample information up until time t. A two-sided filter or smoother estimates the state at time t based on the full sample τ information

which recursively updates the entire path of smoothed states by a magnitude that depends on the one-step-ahead prediction error. The smoothed state estimates at the end of the sample will thus not change if the missing future observations are replaced by the model's own forecasts. Consequently, the closer the predictions of the underlying model beyond the end of the sample to the incoming data, the less severe are the revisions to the past states. Importantly, the end-point problem is not automatically mitigated by the inclusion of exogenous (off-model) forecasts of the observables, but its severity will depend on the ex-post accuracy of these forecasts. This is particularly important around business cycle turning points, where off-model forecasts can be highly imprecise, compounding the end-point problem instead of mitigating it[3].

As pointed out recently by Coibion et al. (2018) revisions to potential output estimates of some institutions tend to be pro-cyclical, that is potential output growth is typically revised upwards when actual GDP growth rates are strong, and downwards when actual growth is perceived as weak. As suggested by Deroose et al. (2019), it is important to highlight the difference between the pro-cyclicality of the 'true' measure of potential output and the pro-cyclicality of the revisions to potential output estimates. While the 'true' level of potential output or its growth rate can legitimately show a degree of pro-cyclicality, e.g. due to labour market hysteresis as in Blanchard and Summers (1987) and its possible reversal as suggested by Yellen (2016), the pro-cyclicality of the revisions to potential output estimates can indicate that the estimation method used is affected by the end-point problem. Importantly, the pro-cyclicality of the path of potential output or its components may - at least partly - explain the missing disinflation puzzle after the financial crisis as in Blanchard et al. (2015), as well as the missing inflation phenomenon in the more recent recovery phase.

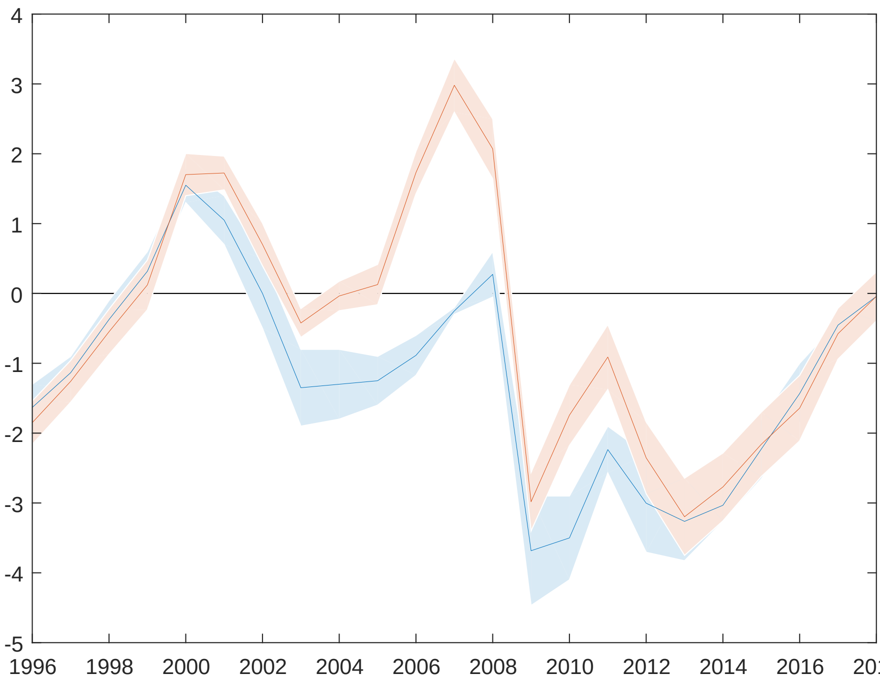
International institutions such as the European Commission, the OECD or the IMF have typically employed production function based methods[4] to estimate the level or growth rate of potential output, see Havik et al. (2014). The production function approach is not a stand-alone method and can cover any of the trend-cycle decomposition methods mentioned above. While projections are usually included in the estimation sample, potential output or output

[3]This is why it is generally a good idea to check the forecasting performance of the underlying models - that are sometimes implicit in the applied filtering method - over an extended horizon with respect to the observables included in the estimation.

[4]The production function approach in its most basic form typically assumes a Cobb-Douglas functional form and yields potential output estimates by independently extracting the trend components of production factors, such as labour, capital and total factor productivity via uni-variate, two-sided filters (or smoothers), for example the Hodrick-Prescott filter. The production inputs are then aggregated based on the assumption that their contributions to output growth equal their respective income shares

gap estimates have often revised over time particularly around turning points in the business cycle. Indeed, real-time output gap vintages of the euro area[5] from these three institutions show consistently lower output gap estimates between 1999 and 2013 than in more recent vintages (Figure 1). The discrepancy is particularly large in the run-up to the 2008 crisis, between the mid-2000s and 2007, where real time estimates indicated a degree of slack or a close to neutral cyclical position, while more recent estimates show massively positive output gaps. Provided that the more recent vintages are better approximations of the 'true' cyclical position in this period, these methods and the projection errors made in the same period contributed to overly optimistic assessments of the state of the euro area economy at a crucial juncture.

Figure 1: Range of real time and latest vintage output gap estimates of the EC, OECD and the IMF for the euro area



Notes: Real time estimates refer to output gaps entailed in each year's autumn forecast exercise rounds. The red band indicates the latest (Autumn 2018) estimates, while the blue band the real time ones.

The current paper introduces a procedure to estimate potential output in a state-space framework with an embedded production function aiming at diminishing the end-point problem. The advantage of this approach is that it combines a growth accounting framework with the joint identification of key trends and cyclical components in a single system of equations.

[5] The relationship between real time and latest vintage estimates are similar to that of a one-sided vs a two-sided filter, with the important caveat that the underlying model can change between vintages.

Consequently the system properties of the model underlying the estimation of potential output can be thoroughly checked. A set of connections between cyclical components ensures that the information content of several macroeconomic time series with respect to the state of the business cycle is exploited in a way that is consistent with some key economic relationships. On top of this the model adds inter-dependencies not only across cyclical components but also across trends. The stochastic trends in the model are disciplined via anchors relating them to observable variables (such as an inflation 'target' in case of trend inflation) and long-run relationships (such as a link between trend wage inflation, trend inflation and trend productivity growth).

The approach introduced in this paper is similar to Proietti et al. (2007) and Proietti and Musso (2012) with a number of key differences. The richer economic structure provides more discipline over key trends in the current model, such as trend productivity growth and allows for the consideration of hysteresis effects. The Bayesian approach applied enables the estimation of the model over a historical data sample without relying heavily on back-data. Jarociński and Lenza (2018) also estimate and evaluate output gaps for the euro area, however neither of their model specifications include a production function - which would allow for assessing the drivers of potential growth - or feature relationships across the trend components.

The paper is organised as follows: Section 2 spells out the structure of the underlying state-space model that will be used to decompose observable variables to trend and cycle components, section 3 describes the data that enter the filtering/smoothing exercise as observations, section 4 provides the details of the Bayesian estimation procedure that is used to parametrise the model and the application of the Kalman-filter (and smoother). Section 5 discusses the key results of the exercise as well as the sources of uncertainty around the estimated states. Section 6 conducts an evaluation exercise focusing on revision properties of the estimated output gap measure and the forecasting performance of the underlying state-space model. Section 7 concludes.

2 The Model

The model introduced in this paper is a state-space model that is built around a production function and a set of equations representing well-known economic relationships. The model does a joint trend-cycle decomposition of six observable variables (real GDP, unemployment rate, core inflation, wage growth, labour force participation rate, hours worked per person) and uses additional two observables (the level of the capital stock and working age population) to directly inform the corresponding trend components. The model features a price- and a wage Phillips curve, as well as an Okun's law type equation that are written in terms of the cyclical component of GDP and the unemployment rate. The trend component of output is decomposed according to a Cobb-Douglas production function. Trend inflation is pinned down by a - possibly time varying - long-term anchor, while trend wage growth is assumed to be the sum of trend inflation and trend (labour) productivity growth. In contrast with more structural approaches such as DSGE models, the equations of the current model are not directly derived from theoretical first principles which admittedly yields a lower degree of internal consistency.

$$Y_t = A_t {L_t}^{\nu} {K_t}^{1-\nu}$$

According to a common specification of the Cobb-Douglas production function observed real GDP (Y_t) can be written in terms of the contribution of production factors, such as labour (L_t), capital (K_t) and a residual term (A_t) often interpreted as total factor productivity (TFP). Under some regularity conditions (e.g. production factors being paid their marginal products) the labour elasticity (ν) of the production function corresponds to the observable labour income share. Taking natural logarithms, the production function takes the following linear form (where lowercase letters denote the natural logarithms of their uppercase counterparts):

$$y_t = a_t + \nu l_t + (1-\nu)k_t$$

Labour input, defined as total hours worked (l_t), can be further decomposed to working age population (wp_t), the labour force participation rate (lp_t), the (un)employment rate (u_t)[6] and hours worked per person or average hours worked (ah_t).

[6]Since the aim is to keep the model linear, for practical purposes instead of $ln(1-u_t)$, the natural logarithm of the employment rate expressed in terms of the unemployment rate, its first order approximation $-u_t$ is used. This allows for the direct estimation of trend unemployment rate or the NAIRU

$$l_t = wp_t + lp_t - u_t + ah_t$$

Assuming that (log) GDP and the (log) production factors can be decomposed into the sum of trend and cyclical components, potential output can be written as the sum of trend TFP and the contributions of trend labour and capital, where the trends of working age population and the capital stock are assumed to equal their observable counterparts. It also follows that the cyclical component of real GDP (or the output gap) can be decomposed as the sum of the TFP cycle and the contribution of the cyclical component of labour input.

$$\bar{y}_t + \hat{y}_t = \bar{a}_t + \hat{a}_t + \nu(\bar{l}_t + \hat{l}_t) + (1-\nu)k_t$$

It is important to stress, that the cyclical and trend component of the level of TFP, $\hat{a}_t$ and $\bar{a}_t$, respectively, are not directly modelled, but can always be recovered by using the accounting identity of the production function and the respective cyclical or trend components that are directly modelled. The cyclical component of TFP, defined as above, can be considered as a proxy for factor utilisation that is not explicitly captured by the model. It may thus implicitly account for cyclical developments in the capital stock and working age population, or changes in labour utilisation that is not captured by adjustment over the typical extensive (employment) or intensive (hours worked per person) margins, see Fernald and Wang (2016).

$$\hat{a}_t = \hat{y}_t - \nu\hat{l}_t \text{ and } \bar{a}_t = \bar{y}_t - \nu\bar{l}_t + (1-\nu)k_t$$

Building on the above considerations, the model can be expressed as the following system of equations, where lower case letters refer to natural logarithms, except for variables explicitly defined in percentage (change) terms.

$$\begin{aligned}
y_t &= \bar{y}_t + \hat{y}_t \\
\hat{y}_t &= \alpha_1 \hat{y}_{t-1} + \alpha_2 \hat{y}_{t-2} + \varepsilon_{\hat{y},t} \\
\bar{y}_t &= \bar{y}_{t-1} + \nu\Delta\bar{l}_t + (1-\nu)\Delta k_t + \widetilde{a}_t \\
\widetilde{a}_t &= \widetilde{a}_{t-1} + \varepsilon_{\widetilde{a},t}
\end{aligned} \tag{1}$$

Observed real GDP (y_t) is decomposed into a trend ($\bar{y}_t$) and a cyclical component ($\hat{y}_t$), the output gap. To allow for richer cyclical dynamics, the output gap is assumed to follow an AR(2) process. Trend output is determined by a Cobb-Douglas type production function, where $\bar{l}_t$ is trend total hours worked, k_t is the capital stock, $\tilde{a}$ is trend TFP growth, ν corresponds to the labour income share and Δ is a first difference operator. Trend TFP growth ($\widetilde{a}$), is assumed to follow a random walk, thus the level of TFP is I(2), or an integrated random walk[7]. Trend labour input, $\bar{l}_t$ (total hours), can be broken down to $wp_t + \overline{lp}_t - \bar{u}_t + \overline{ah}_t$, where wp_t is working age population, $\overline{lp}_t$ is the trend of the participation rate, $-\bar{u}_t$ is a linear approximation of the trend employment rate represented in terms of the trend unemployment rate (or NAIRU), while $\overline{ah}_t$ is the trend component of average hours worked.

$$\begin{aligned}
lp_t &= \bar{lp}_t + \hat{lp}_t \\
\widehat{lp}_t &= -\gamma_4 \hat{u}_{t-1} + \varepsilon_{\widehat{lp},t} \\
\overline{lp}_t &= \overline{lp}_{t-1} + \widetilde{lp}_{t-1} \\
\widetilde{lp}_t &= \widetilde{lp}_{t-1} + \varepsilon_{\widetilde{lp},t}
\end{aligned} \tag{2}$$

The trend-cycle decomposition of the labour force participation rate (lp_t) entails a trend component that follows an integrated random walk. It is assumed that the cyclical component is related to the unemployment gap, capturing possible discouraged (or added-) worker effects.

$$\begin{aligned}
u_t &= \bar{u}_t + \hat{u}_t \\
\hat{u}_t &= \gamma_1 \hat{u}_{t-1} - \gamma_2 \hat{y}_{t-1} + \varepsilon_{\hat{u},t} \\
\bar{u}_t &= \bar{u}_{t-1} + \kappa_1 \Delta \hat{u}_t + \varepsilon_{\tilde{u},t}
\end{aligned} \tag{3}$$

The cyclical component of the unemployment rate ($\hat{u}_t$) is subject to an Okun's law type relationship, i.e., the unemployment gap is connected to the output gap. As shown by Christiano et al. (2020) an Okun's relationship written in terms of the output and unemployment gaps can be derived from a slightly modified, basic new Keynesian DSGE model. The trend unemployment rate or NAIRU ($\bar{u}_t$) is assumed to follow a random walk process with a drift term, where the latter captures changes in the unemployment gap, enabling hysteresis-type effects. This modelling choice reflects the observation that temporary shocks to (un)employment can have persistent effects due to skill-erosion, see Pissarides (1992). It also implies that possible hysteresis effects

[7] While the notation could be further simplified, its current form highlights the similarities to the local linear trend model, as applied e.g. by Clark (1987)

are symmetric, i.e. not confined to economic busts but can occur in booms as well, as suggested by Yellen (2016). For example, in the mature phase of the business cycle emerging labour shortages can make firms more willing to hire unemployed with low-employability characteristics and upgrade their skills via e.g on-site training. For a somewhat different implementation of labour market hysteresis in a multivariate filter framework see Alichi et al. (2019).

$$\begin{aligned} ah_t &= \bar{ah}_t + \hat{ah}_t \\ \hat{ah}_t &= \gamma_6 \hat{y}_t + \varepsilon_{\hat{ah},t} \\ \bar{ah}_t &= \bar{ah}_{t-1} + \widetilde{ah}_{t-1} \\ \widetilde{ah}_t &= -\kappa_2 \widetilde{lp}_{t-1} + \varepsilon_{\tilde{ah},t} \end{aligned} \tag{4}$$

Hours worked per person (or average hours, ah_t) is decomposed similarly to the labour force participation rate. The cyclical component ($\hat{ah}_t$) is assumed to be driven by the contemporaneous output gap reflecting that hours worked per person is typically the first margin of adjustment to shocks affecting the labour market, in particular in the euro area. In the absence of regulatory changes, it is assumed that changes in trend hours worked per person ($\bar{ah}_t$) are due to changes in the labour market participation of demographic groups with preferences towards part-time work, such as female or elder workers, see e.g. Bodnár (2018). This is consistent with the presence of the growth rate of the trend participation rate in the trend average hours growth ($\widetilde{ah}_t$) equation.

$$\begin{aligned} wp_t &= \bar{wp}_t \\ \bar{wp}_t &= \bar{wp}_{t-1} + \widetilde{wp}_{t-1} \\ \widetilde{wp}_t &= \widetilde{wp}_{t-1} + \varepsilon_{\tilde{wp},t} \end{aligned} \tag{5}$$

$$\begin{aligned} k_t &= \bar{k}_t \\ \bar{k}_t &= \bar{k}_{t-1} + \widetilde{k}_{t-1} \\ \widetilde{k}_t &= \widetilde{k}_{t-1} + \varepsilon_{\tilde{k},t} \end{aligned} \tag{6}$$

Developments in working age population (wp_t) and the capital stock ($\bar{k}_t$) are assumed to be observable and are not decomposed to trend and cycle within the model, which is standard in the literature. The applied modelling framework requires however to define their unobservable counterparts that enter the state equations. The latter are modelled as integrated random walk processes and are directly linked to their observable counterparts.

The model is rounded up by connections between the real and nominal side. A measure of (core) inflation is decomposed to trend and cyclical components, where the latter is assumed to have some intrinsic persistence, i.e. a dependence on past cyclical inflation due e.g. to the price-setting mechanism as described in Angeloni et al. (2006). The cyclical component of inflation is connected to the output gap via a Phillips curve-type relationship.

$$\begin{aligned} \pi_t &= \bar{\pi}_t + \hat{\pi}_t \\ \hat{\pi}_t &= \beta_1 \hat{\pi}_{t-1} + \beta_2 \hat{y}_{t-1} + \varepsilon_{\hat{\pi},t} \\ \bar{\pi}_t &= (1-\phi)\pi^* + \phi \bar{\pi}_{t-1} + \varepsilon_{\bar{\pi},t} \end{aligned} \tag{7}$$

where π_t is a measure of core inflation, $\bar{\pi}_t$ is trend inflation, $\hat{\pi}_t$ is the inflation gap and π^* is an inflation anchor. Introducing trend and cyclical components, as e.g. in Stock and Watson (2007), allows for separating the impact of temporary and permanent shocks to inflation where the latter may capture developments possibly related to structural factors, such as globalisation or (unobservable) inflation expectations that are relevant for price setting. From a different angle, trend inflation can be thought of as a time-varying intercept term in a traditional Phillips curve setup. If $\phi < 1$ core inflation is backward looking, but is anchored by a constant. Setting $\phi = 1$ yields a random walk trend which allows for shocks with permanent effects on inflation.

$$\begin{aligned} w_t &= \bar{w}_t + \hat{w}_t \\ \hat{w}_t &= \beta_3 \hat{w}_{t-1} + \beta_4 \hat{u}_{t-1} + \varepsilon_{\hat{w},t} \\ \bar{w}_t &= \bar{\pi}_t + \Delta \bar{y}_t - \Delta \bar{l}_t + \varepsilon_{\bar{w},t} \end{aligned} \tag{8}$$

To connect the labour market cycle with wage growth, a wage Phillips-curve is added to the model. The cyclical component of wage growth ($\hat{w}_t$) has an AR(1) term reflecting intrinsic persistence, e.g. due to wage rigidity or indexation. Furthermore, the wage growth cycle is partly explained by the unemployment gap while trend wage growth ($\overline{w}_t$) is assumed to be the sum of trend inflation and trend labour productivity growth ($\Delta \bar{y}_t - \Delta \bar{l}_t$)[8], capturing a long run relationship connecting these variables. The trend wage growth equation implies that over the long run real wage growth equals labour productivity growth, in other words the trend growth rate of real unit labour costs is zero. This equation is key for disciplining the behaviour of trends in the model since it connects all of them.

[8]if wage growth and inflation is defined in annualised quarterly growth rates this term needs to be multiplied by 4

Given the Cobb-Douglas production function, trend labour productivity growth can be further decomposed to trend TFP growth and capital deepening:

$$\Delta\bar{y}_t - \Delta\bar{l}_t = \widetilde{a}_t + (1-\nu)(\tilde{k}_t - \Delta\bar{l}_t).$$

The disturbance terms in the state equations of the model are assumed to have a mean zero Gaussian distribution with a diagonal covariance matrix, $\mathbf{\Sigma}_\epsilon$, with σ_i^2 denoting the variance of the i-th state disturbance. In the following sections two different specifications of the above model are reported, reflecting modelling choices related to trend inflation (Table 1). In terms of modelling trend inflation (7) one specification features an AR(1) process with a constant mean (π^*) set to a value consistent with the ECB's inflation objective, adjusted for the difference in the historical averages of headline and core inflation (model A). The other specification has a stochastic trend ($\phi = 1$), allowing for permanent shocks affecting trend inflation (model B). In the rest of the paper, model A will be considered as the baseline specification.

Table 1: Model versions

Feature	Model A	Model B
ϕ	< 1	$= 1$
π^*	target equivalent	n.a.

3 Data

The model introduced in Section 2 is estimated on publicly available data collected from Eurostat and the European Commission's AMECO database. The estimation sample is constrained by the availability of national accounts data for the euro area aggregate which typically starts in 1995 Q1. While the common currency was non-existent before 1999, it can be reasonably assumed that starting from the mid-1990s, after the ratification of the Maastricht treaty, expectations and policies became increasingly aligned with the approaching formation of the monetary union. On the other hand, backdated euro area data going back far before 1995 include periods in which future EA countries had more severe heterogeneity in terms of nominal developments and economic structures (e.g. West vs. East Germany). On top of this, the time series properties of a purely backdated subsample would reflect that of the backdating method and not necessarily the data generating processes. Due to the above issues the estimation of the model is conducted on actual historical euro area data that is typically available from 1995, and backdating is used only when a time-series has missing values at the beginning of the sample. The source of the backdated data is the Area-Wide Model database, see Fagan et al. (2001).

The model is estimated on aggregate euro area data at a quarterly frequency. The output measure in the model corresponds to the volume of chain linked real GDP. Core inflation is HICP inflation excluding food and energy, measured in terms of annualised quarter-on-quarter (q-o-q) growth rates. Wage growth corresponds to the annualised q-o-q growth rate of hourly compensation. The unemployment rate is the harmonised unemployment rate as defined by Eurostat, the ratio of the number of unemployed workers to the labour force. The participation rate is the ratio of the labour force (i.e. the number of people who are in employment or unemployed) and the working age population, defined as the population at least 15 years old but younger than 75 years.

All quarterly series are seasonally and working day adjusted by the disseminating authority, except for HICP excluding food and energy which is adjusted with the US Census Bureau's x13-ARIMA package. Data on the capital stock and working age population are from the European Commission's Annual Macro-economic (AMECO) database and interpolated to quarterly frequency using a piecewise cubic Hermite interpolation method, to minimise spurious oscillations. Since the latter two variables are slowly moving and less volatile time series, interpolation is not expected to introduce any significant distortions. Table 2 summarises the data set and the applied transformations.

Table 2: Data description

Alias	**Description**	**Freq.**	**Range**	**Source**	**Transformations**
y	Real GDP, chain linked, bln euros	Q	1995Q1-2019Q3	Eurostat	natural logarithm
π	HICP excluding food and energy, index	Q	1996Q1-2019Q3	Eurostat	nat. log., seasonal adjustment, 4*first difference
w	Hourly compensation, index	Q	1995Q1-2019Q3	Eurostat	nat. log., 4*first difference
u	Unemployment rate, Eurostat definition	Q	1997Q1-2019Q3	Eurostat	divided by 100
ah	Hours worked per employed person	Q	1995Q1-2019Q3	Eurostat	nat. log.
lp	Labour force per working age population (15-74 yrs)	Q,A	1997Q1-2019Q3	Eurostat, AMECO	nat. log.
k	Net capital stock	A	1995-2019	AMECO	nat. log., interpolated to quarterly freq.
wp	Working age populataion (15-74 yrs)	A	1995-2019	AMECO	nat. log., interpolated to quarterly freq

4 Estimation strategy

The unobserved states and the parameters of the model are both estimated with the support of the Kalman filter, a recursive algorithm that can be applied in a straightforward way to the state-space representation of the model. Denoting by $\boldsymbol{\alpha}_t$ the vector of unobservable states, $\boldsymbol{y_t}$ the vector of measurement variables, $\boldsymbol{\epsilon}_t$ the shock vector of state equations with normally distributed mean zero shocks and covariance matrix $\boldsymbol{\Sigma}_\epsilon$ and by $\boldsymbol{\eta}_t$ the measurement shock vector with covariance matrix $\boldsymbol{\Sigma}_\eta$, the general state space form can be written as follows:

$$\boldsymbol{\alpha}_t = \boldsymbol{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{K} + \boldsymbol{R}\boldsymbol{\epsilon}_t$$
$$\boldsymbol{y}_t = \boldsymbol{Z}\boldsymbol{\alpha}_t + \boldsymbol{H}\boldsymbol{\eta}_t$$

$$\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon)$$
$$\boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta)$$

Given measurements $\boldsymbol{y_t}$, the initial state vector $\boldsymbol{\alpha}_0$ and initial state covariance $\boldsymbol{P_0}$ and matrices $\boldsymbol{T}$, $\boldsymbol{K}$, $\boldsymbol{R}$, $\boldsymbol{Z}$, $\boldsymbol{H}$, $\boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{\Sigma}_\eta$ the Kalman filter provides optimal (in terms of mean squared error) linear estimates of the unobserved states and their covariance (see Appendix A for the full state-space representation of the model). In case of a linear model and Gaussian disturbances the state estimates will also have a multivariate Gaussian distributions. The Kalman filter provides state estimates for any time t, based on information available up to time t. However, in economic applications such as this paper, one typically wants to consider the full sample (τ) information when inferring the state of the economy at time $t \leq \tau$. Therefore, the state estimates reported in this paper refer to smoothed - or two-sided - estimates derived using a fixed interval smoother algorithm that is directly related to the Kalman filter (Anderson and Moore, 1979, Ch. 7.4).

To be able to estimate the states, the Kalman filter needs to be initialised via assigning values to $\boldsymbol{\alpha}_0$ and $\boldsymbol{P_0}$. For stationary state variables, such as the cyclical components, the Kalman filter can be initialised at their model-implied distribution, i.e. their unconditional mean and state covariance. In case of the state variables that are modelled as non-stationary stochastic processes the initialisation of the Kalman filter is less straightforward. One way to initialise non-stationary series is to apply a diffuse prior, with zero mean and infinite (in practice a very large, but finite) variance. In the current case a diffuse initialisation would entail excessively large initial uncertainty, given that the initial non-stationary states cannot be too far from their observable counterparts at the beginning of the sample or in the last pre-sample periods. For example, it would be unreasonable to admit the possibility that the level of potential GDP or the NAIRU were 0 - or even negative - in the last pre-sample period. An alternative is to initialise the Kalman filter with a vector of fixed, but unknown scalars that are treated as parameters and estimated in a likelihood maximising way utilising the one-step ahead prediction errors with respect to the first observations in the sample as described e.g. in Harvey (1990), section 3.4.4. Since this approach uses sample information, it is controversial in a Bayesian estimation context. A third alternative is to initialise the non-stationary states with a proper prior distribution, that can be based e.g. on some pre-sample information. The results reported in the following sections the initialisation of the Kalman filter will be based on the third approach. In particular, the stationary states are initialised at zero, while the ones defined as non-stationary are assumed to correspond to the last pre-sample values of their observable counterparts, taken (and interpolated to quarterly frequency) from the AMECO database. The initial state covariance matrix is assumed to be diagonal, with the diagonal elements corresponding to approximately 5 percentage (points) standard deviations (see Appendix A).

Besides the unobserved states and their covariance, the parameters of the model can also be estimated with the help of the Kalman-filter. It is well known that under some regularity conditions the one-step ahead prediction errors and their covariance provided by the Kalman-filter can be used to calculate the sample likelihood function. This enables the application of numerical optimisation methods to find the extrema of the (log-) likelihood function in case of more complex problems, where a closed form solution is not available. Bayesian methods that combine prior beliefs and the information in the data (the likelihood function) to learn about model parameters ($\boldsymbol{\theta}$) can be particularly useful when the number of parameters to be estimated is large but the size of the sample is relatively small. Furthermore, Bayesian methods can help to overcome identification problems, i.e. the likelihood surface being flat or irregular in certain dimensions, at the cost of having informative priors which can make the results sensitive to the choice of prior distributions. Due these considerations, the model parameters are estimated with Bayesian methods[9].

Applying Bayes' theorem and assuming away the normalising constant that does not depend on $\boldsymbol{\theta}$, the log of the posterior distribution, $F(\boldsymbol{\theta})$, can be expressed as the sum of the log likelihood of the data and the log prior:

[9]The model is estimated with the help of the Iris toolbox for Matlab, see Benes et al. (2020)

$$F(\boldsymbol{\theta}) = \sum_{t=1}^{\tau} \ln P(\boldsymbol{y_t} \mid \boldsymbol{\theta}) + \ln P(\boldsymbol{\theta})$$

In the current exercise four basic types of prior distributions are used for the different parameters of the model (see Table 3). For the coefficients of auto-regressive processes of order one, or AR(1), a beta distribution is assumed, which has a support (0,1). This is consistent with the prior belief that the cyclical components of the observable input series are stationary. The beta priors on the AR(1) coefficients have a mean of 0.7 and a standard deviation of 0.15. These hyper-parameters imply fairly persistent but stationary processes that are clearly distinguishable from I(1) processes, since close to 90 per cent of the probability mass of these prior distributions fall between 0.5 and 0.95. In case of the output gap that is defined as an AR(2) process as in Clark (1987), the above beta prior is used on the sum of the autoregressive coefficients, following the procedure of imposing system priors by Andrle and Plašil (2018). Given the (0,1) support of the beta distribution, this is sufficient to ensure stationarity of the output gap.

Most of the other coefficients of the model equations are assumed to have gamma-type priors that have a support of $(0, \infty)$. These reflect prior beliefs in the existence and the direction of the economic relationships these parameters capture. For example, in the price and wage Phillips curves it is a non-controversial assumption that inflation and wage growth are pro-cyclical - on the basis of economic theory. Similarly, in case of the Okun's law the relationship between the output gap and the unemployment gap is assumed to be strictly negative. For these coefficients the literature often finds values that are between zero and 1 and typically closer to zero than one. This is reflected in the choice of hyper-parameters that imply that more than 90 per cent of the probability mass lies between 0 and 1, and close to 60 percent of it between 0 and 0.5. Some coefficients capturing less well established relationships, such as the hysteresis term in the trend unemployment rate equation have Gaussian prior distributions centred around zero. For numerical stability reasons, some of the prior densities are truncated in regions with very small probability mass, see Table 3.

All estimated shock standard deviations in the state equations of the model are assumed to have an inverse gamma (IG) prior distribution with means corresponding to 1 and 0.01 in case of the cyclical and trend shocks, respectively. The IG distribution cannot take the value of zero, thus its application as a prior on shock standard deviations helps to alleviate the so-called pile-up problem, as described by Stock (1994). The standard deviations of all inverse gamma priors are set to ∞, which is common in the literature, see e.g. in Adolfson et al. (2007). It is important to note that in case of the IG distribution an infinite standard error still yields a proper prior distribution, with finite values of the distribution's scale and shape parameters. The labour share coefficient in the production function (ν) and the shock standard deviations in the capital stock and working age population equations ($\sigma_{\tilde{k}}$ and $\sigma_{\tilde{wp}}$, respectively) are not estimated. The labour share is set to its long term average value, 0.63. The shock standard deviations of the capital stock and working age population are fixed at values corresponding to the prior mean of the estimated trend shock standard deviations (0.01).

Parameter estimation is conducted with the help of a fairly standard Bayesian procedure following a the algorithm described e.g. by Herbst and Schorfheide (2016). First the mode ($\boldsymbol{\theta}^*$) of the log-posterior, $F(\boldsymbol{\theta})$, is found via a numerical optimisation method where the likelihood is evaluated in each step with the help of the Kalman-filter. While the numerical optimisation provides the posterior modes - possibly local maxima - and the corresponding Hessian, proper Bayesian inference requires the full posterior distribution. Given the type of prior densities used, the posterior distribution cannot be derived analytically, neither are the conditional posteriors available, thus direct sampling from them is infeasible. In such cases the posterior distribution can be simulated using a Markov chain Monte Carlo (MCMC) method based on Metropolis et al. (1953). In particular, the posterior simulation in this exercise is conducted with the help of an adaptive random walk Metropolis algorithm, as proposed by Haario et al. (1999). The algorithm is initialised with $\boldsymbol{\theta}^*$ and the corresponding covariance matrix, which is the inverse of the negative of the Hessian evaluated at the mode, and uses a multivariate normal proposal distribution centered around zero to perturb the current state of the chain. Draws from the proposal distribution are added to the current state of the chain to form a candidate state. The log posterior of the candidate state (i.e. the log-likelihood plus the log-prior of the model parameters) is then evaluated, with the help of the Kalman-filter. If the posterior evaluated at the candidate state is higher than or equal to the actual state of the chain it is automatically accepted, otherwise it is accepted with a probability that equals the ratio of the posteriors evaluated at the current and the candidate state. This allows the chain to explore the lower probability regions of the posterior density. To increase the efficiency of the algorithm the covariance matrix of the proposal distribution is adapted in each draw to achieve a target acceptance ratio of 23.4 per cent, a standard value in the literature, see e.g. Chapter 7.8.4 of

Robert and Casella (2005). Since the adaptation satisfies the diminishing adaptation condition[10], the simulated Markov chains will retain the posterior density as their stationary distribution, as shown in Chapter 4.3.1 of Brooks et al. (2011). The posterior chains are simulated in a parallel setup resulting in 16 chains, each containing 200,000 draws after discarding an initial 200,000 draws. The (marginal) posterior distributions of model parameters, as well as trace plots of the parallel chains are reported in Appendix B allowing for a visual inspection of the convergence of the chains to their stationary distribution. The multivariate convergence diagnostics of Brooks and Gelman (1998) which can be used to assess the convergence of the parallel chains numerically, yields a potential scale reduction factor of 1.01, suggesting marginal gains if the number of draws were to be increased further. Since the evidence suggests that the chains have converged to their stationary distributions, the parallel draws are pooled together and the posterior statistics are calculated over 3.200.000 draws for each parameter.

Table 3: Prior and posterior statistics (model A)

	Prior					Posterior			
Parameter	type	l.b.	u.b.	mean	std	mean	median	90% HPDI	
$\alpha_1 + \alpha_2$	beta	0	1	0.7	0.15	0.930	0.939	0.864	0.999
β_1	beta	0	1	0.7	0.15	0.588	0.59	0.361	0.824
β_2	gamma	0	10	0.5	0.3	0.109	0.101	0.026	0.189
β_3	beta	0	1	0.7	0.15	0.469	0.466	0.259	0.682
β_4	gamma	0	10	0.5	0.3	0.324	0.306	0.089	0.547
γ_1	beta	0	1	0.7	0.15	0.804	0.819	0.652	0.96
$-\gamma_2$	gamma	0	10	0.5	0.3	0.122	0.113	0.026	0.21
$-\gamma_4$	normal	-10	10	0	1	0.125	0.126	-0.334	0.575
γ_6	normal	-10	10	0	1	0.199	0.196	0.03	0.37
κ_1	normal	-10	10	0	1	0.259	0.257	-0.095	0.613
κ_2	normal	-10	10	0	1	0.531	0.534	0.282	0.786
ϕ	beta	0	1	0.7	0.15	0.654	0.666	0.425	0.891
$\sigma_{\hat{y}}$	inv. gam.	0	∞	1	∞	0.012	0.012	0.01	0.014
$\sigma_{\tilde{a}}$	inv. gam.	0	∞	0.01	∞	0.001	0.001	0.001	0.001
$\sigma_{\hat{u}}$	inv. gam.	0	∞	1	∞	0.011	0.011	0.009	0.013
$\sigma_{\tilde{u}}$	inv. gam.	0	∞	0.01	∞	0.002	0.002	0.001	0.002
$\sigma_{\widehat{lp}}$	inv. gam.	0	∞	1	∞	0.011	0.011	0.009	0.013
$\sigma_{\widetilde{lp}}$	inv. gam.	0	∞	0.01	∞	0.001	0.001	0.001	0.001
$\sigma_{\widehat{ah}}$	inv. gam.	0	∞	1	∞	0.011	0.011	0.009	0.013
$\sigma_{\widetilde{ah}}$	inv. gam.	0	∞	0.01	∞	0.002	0.002	0.001	0.002
$\sigma_{\hat{\pi}}$	inv. gam.	0	∞	1	∞	0.012	0.012	0.01	0.014
$\sigma_{\bar{\pi}}$	inv. gam.	0	∞	0.01	∞	0.002	0.002	0.001	0.003
$\sigma_{\hat{w}}$	inv. gam.	0	∞	1	∞	0.017	0.017	0.014	0.02
$\sigma_{\bar{w}}$	inv. gam.	0	∞	0.01	∞	0.003	0.003	0.001	0.004

Note: l.b.: lower bound, u.b.: upper bound, HPDI: highest posterior density interval, i.e. the narrowest interval containing the specified probability mass

[10] In particular, the adaptation is adjusted by a rate of decay of 0.8 with each draw.

Overall, given the a priori beliefs entailed in the prior distributions, the data seems to contain information about most model parameters, i.e. the marginal posteriors do not coincide with the priors. (see Table 3 and Figures B4-B6 in the Appendix). In particular, there is information in the data about the coefficients in Okun's law and the Phillips curves. The data are also informative about the coefficients with prior distributions centered around zero. For example, the data is informative about the coefficient linking the unemployment gap and the NAIRU (κ_1), providing some evidence on two-way labour market hysteresis effects. Similarly, the data suggests a negative relationship between the trend growth rate of the participation rate and that of average hours worked (κ_2), which is consistent with the observation that increases in the labour force participation rate over the last decade have been driven by demographic groups that are more likely to work part-time. A negative relationship seems to connect the cyclical component of the participation rate and the unemployment gap (γ_4) - although with zero falling well within the 90% highest posterior density interval - suggesting that discouraged worker effects play a role over the business cycle. The data suggest a positive relationship between the cyclical component of averages hours worked and the output gap (γ_6), implying that the intensive margin is a potentially important labour market adjustment channel in response to business cycle fluctuations in the euro area.

5 An overview of the estimated states

This section presents the key results of the exercise, in particular the smoothed estimates of the unobservable states and various decompositions. The smoothed state estimates shown below are calculated conditional on the estimated model parameters being at the median of their posterior distributions[11]. Draws from the full conditional distribution of the states over time are shown in section 5.1 and are available from the simulation smoother of Durbin and Koopman (2002).

The estimated trends and the six observables that were made subject to trend-cycle decomposition in the model are shown in Figure 2 (model A) and Figure 4 (model B). While both model specifications feature a break in the slope of potential GDP after the great financial crisis, the level of potential GDP at the end of the sample is estimated to be higher in case of model A. Both specifications feature a similar, counter-cyclical profile of the trend unemployment rate (or NAIRU) that has been declining since 2013, with the estimate of model A being consistently below that of model B. The increasing participation rate trend and the declining path of trend average hours work is consistent with a long-run increase in the labour market participation of demographic groups that prefer to work part-time in both specifications. The trend inflation trajectories reflect the different assumptions regarding the stochastic process governing the trend. The estimated trend of core inflation is flat in case of model A, which assumes a fixed anchor, but it shows a persistent decline starting from 2013 if the trend is specified as a random walk. This is broadly reflected in the trend wage growth trajectories as well. If the possibility of a persistent decline in trend inflation after the financial crisis is ruled out (model A), the weakness in core inflation translates to a lower output gap via the price Phillips curve but also to a somewhat lower rate of potential growth via the trend wage growth equation (8) compared to the stochastic trend inflation specification of model B. This is particularly true in periods where the difference between the two trend inflation measures is widening

Based on data available up to the third quarter of 2019, the baseline specification (model A) suggests a euro area output gap that was in slack territory for 35 consecutive quarters after the financial crisis of 2008/2009, became broadly neutral at the end of 2017 and remained so since then (Figure 3a). In contrast, model B suggests a more complete recovery after the financial crisis, with the output gap closing in the first half of 2011 and rising clearly above zero in 2017 (Figure 5a). Before the financial crisis the output gap is estimated to have peaked between 3

[11] In the case of asymmetric distributions, such as some of the marginal posteriors in this exercise, the median can be a better representative of the density than its mean

and 4 per cent of potential GDP. The key drivers of the output gap in both specification are the TFP cycle and the unemployment cycle, with the the cycle in hours worked per person and the participation rate gap playing a more muted role.

Potential growth of the euro area is estimated to fall between 1.3 and 1.5 per cent in year-on-year terms at the end of 2019. Potential growth has declined over the estimation sample, with the decline starting well before the financial crisis and the ensuing great recession. At the same, time potential growth shows a somewhat pro-cyclical pattern. As the growth-accounting decompositions (Figures 3b and 5b) indicate this pro-cyclicality comes from a few different sources. First, due to the way possible hysteresis effects are captured in the model, i.e. via a link between the cyclical component of the unemployment rate and its trend. Second, a degree of pro-cyclicality appears due the contribution of capital stock changes and working age population growth that enter the production function directly, that is without any trend-cycle decomposition. In case of the former the pre-crisis boom may have led to strong capital-accumulation, particularly in terms of residential properties, that was partly mis-allocated, as suggested e.g. by the narrowing of trend-TFP growth contribution over the same period. In case of working age population growth some volatility may have arisen due to a degree of pro-cyclically in net migration developments in some time periods. Regarding the other potential growth components, the growth rate of trend participation rate is somewhat pro-cyclical, trend hours worked per person a-cyclical while trend-TFP growth is counter-cyclical, based on pairwise correlations with the output gap (Table 4). The GDP growth decompositions (Figures 3c and 5c) shows how changes in its cyclical and trend components contributed to overall GDP growth.

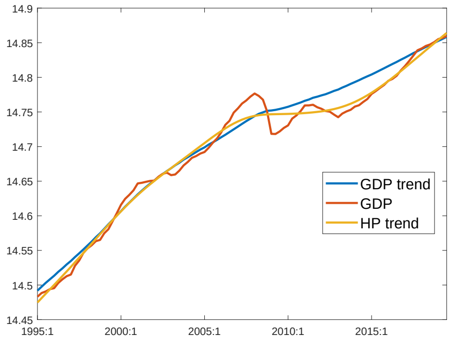
Table 4: Correlation of trend growth components with the output gap

trend growth component	corr. with $\hat{y}$
Δwp	0.36
$\Delta \bar{lp}$	0.32
$\Delta \bar{u}$	-0.35
$\Delta \bar{ah}$	-0.04
Δk	0.68
$\tilde{a}$	-0.20

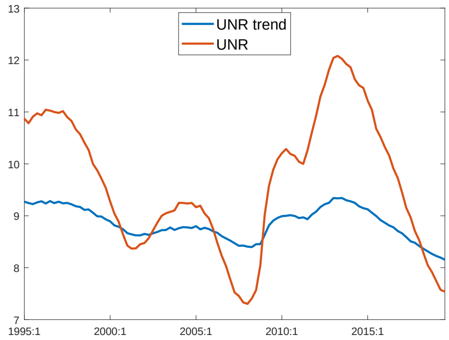
The decomposition of the growth rate of hourly wages (Figure 3d) highlights how the wage growth trend bridges the real and nominal trends of the model. It indicates that the slowdown in wage growth after the great financial crisis was mostly driven by transitory factors, while trend labour productivity (output per hour) growth, decomposed further to capital deepening and trend TFP growth, played a smaller role. If trend-inflation is allowed to follow a stochastic trend (model B), the slowdown in wage growth is explained also by the decline in trend inflation with the contribution from transitory factors shrinking (see 5d). Given observed wage growth and the fact that higher trend inflation is not fully offset by the increased amount of labour market slack in model A, trend productivity growth needs to shift downwards compared to the stochastic trend specification of core inflation of model B.

Figure 2: Key variables and their trends as estimated in Model A

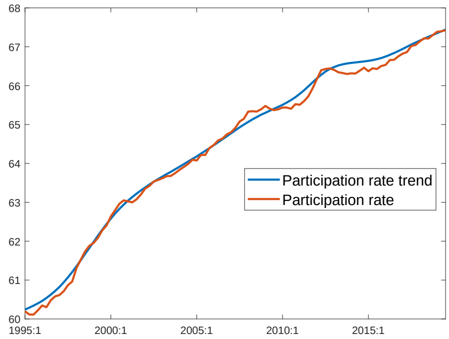
(a) GDP



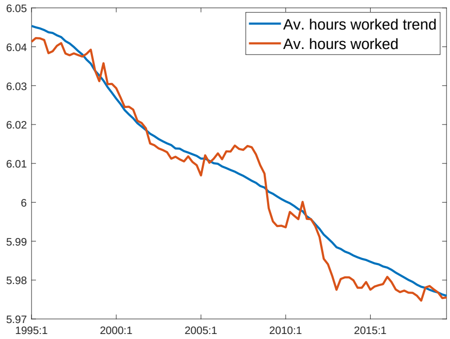
(b) Unemployment rate



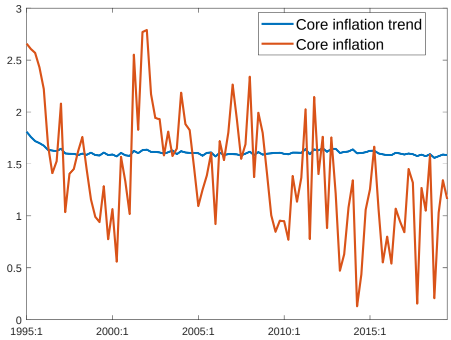
(c) Participation rate



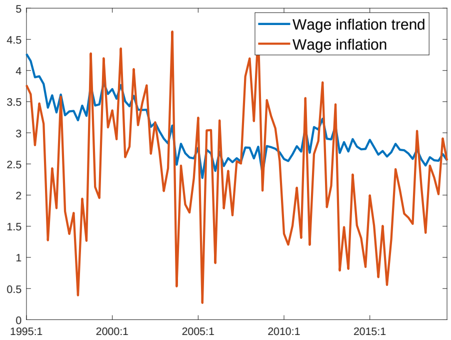
(d) Average hours worked



(e) Core inflation

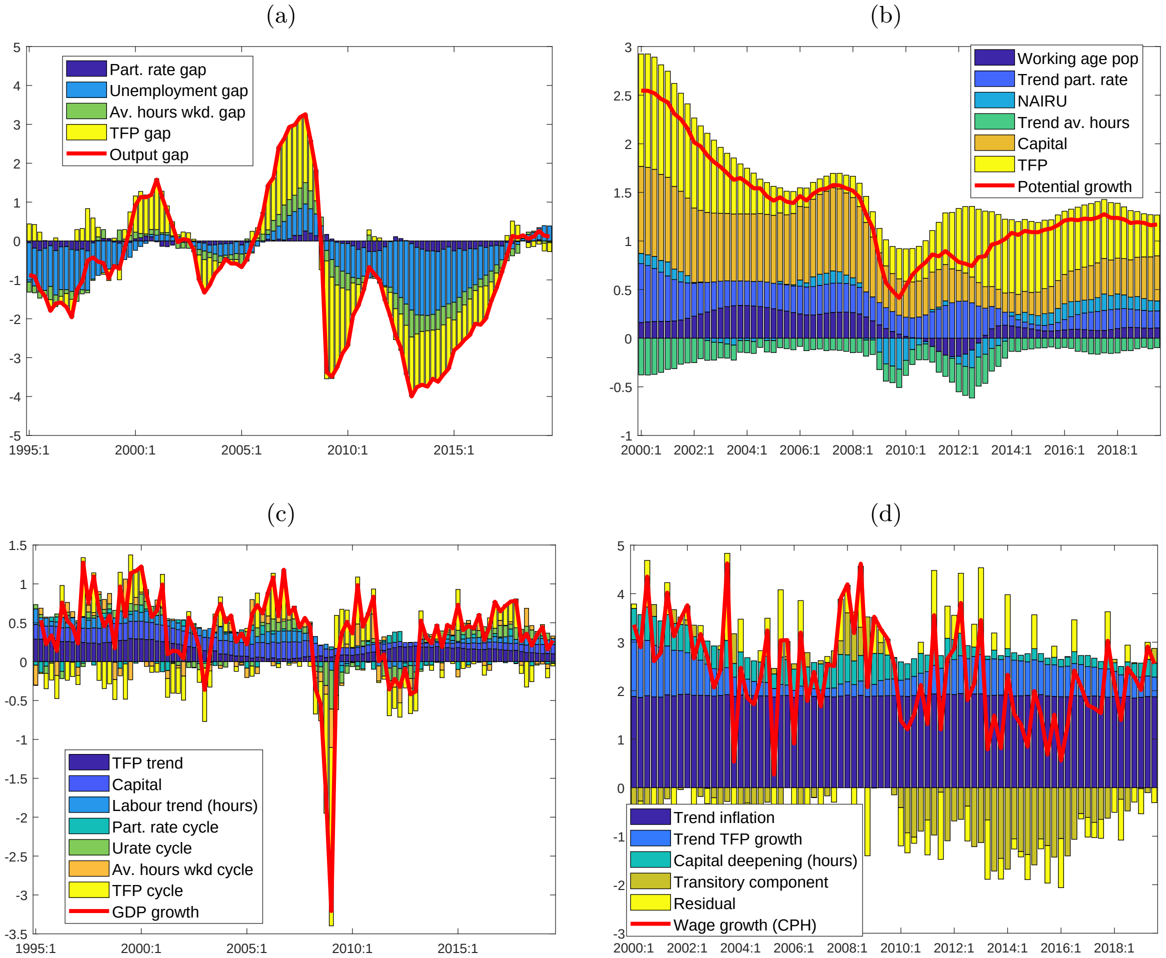


(f) Wage growth



Notes: GDP and hours worked per person are shown in natural logarithm, HP trend is a Hodrick-Prescott smoothed trend with a signal-to-noise ratio of 1/1600; unemployment rate (UNR) and the participation are reported in percentages of the labour force and working age population, respectively; core inflation (HICP excluding food and energy) and wage growth (hourly compensation) are annualised q-o-q growth rates multiplied by 100

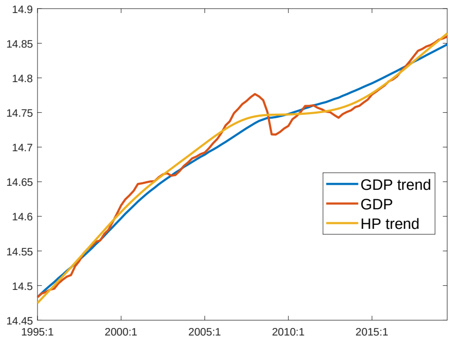
Figure 3: Four key decompositions in Model A



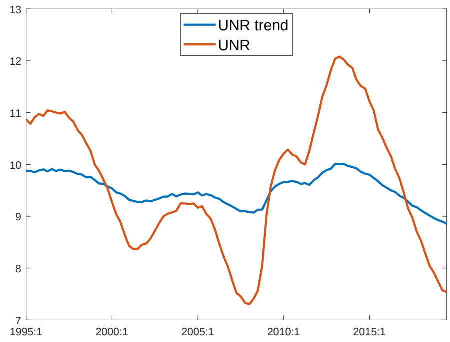
Notes: the output gap is the difference between the natural logarithm of GDP and that of potential output; potential growth is shown in y-o-y growth rates; GDP growth is shown in q-o-q log differences; wage growth is the annualised q-o-q growth rate of hourly compensation. All measures are multiplied by 100

Figure 4: Key variables and their trends as estimated in Model B

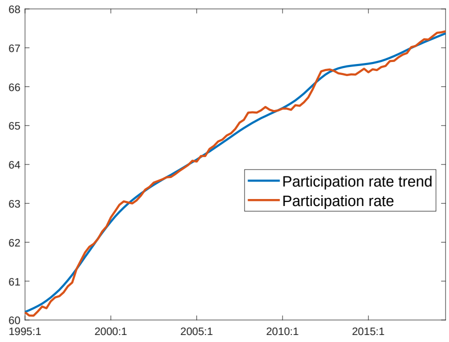
(a) GDP



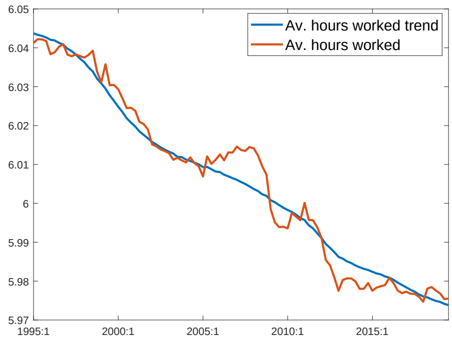
(b) Unemployment rate



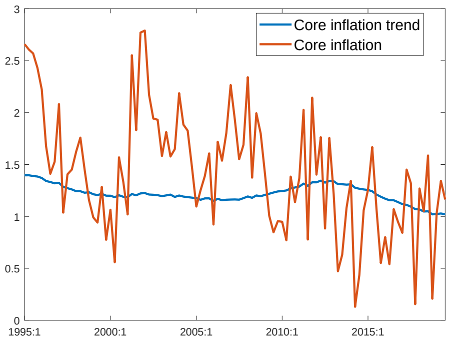
(c) Participation rate



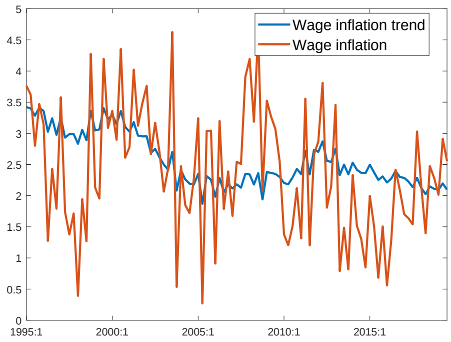
(d) Average hours worked



(e) Core inflation

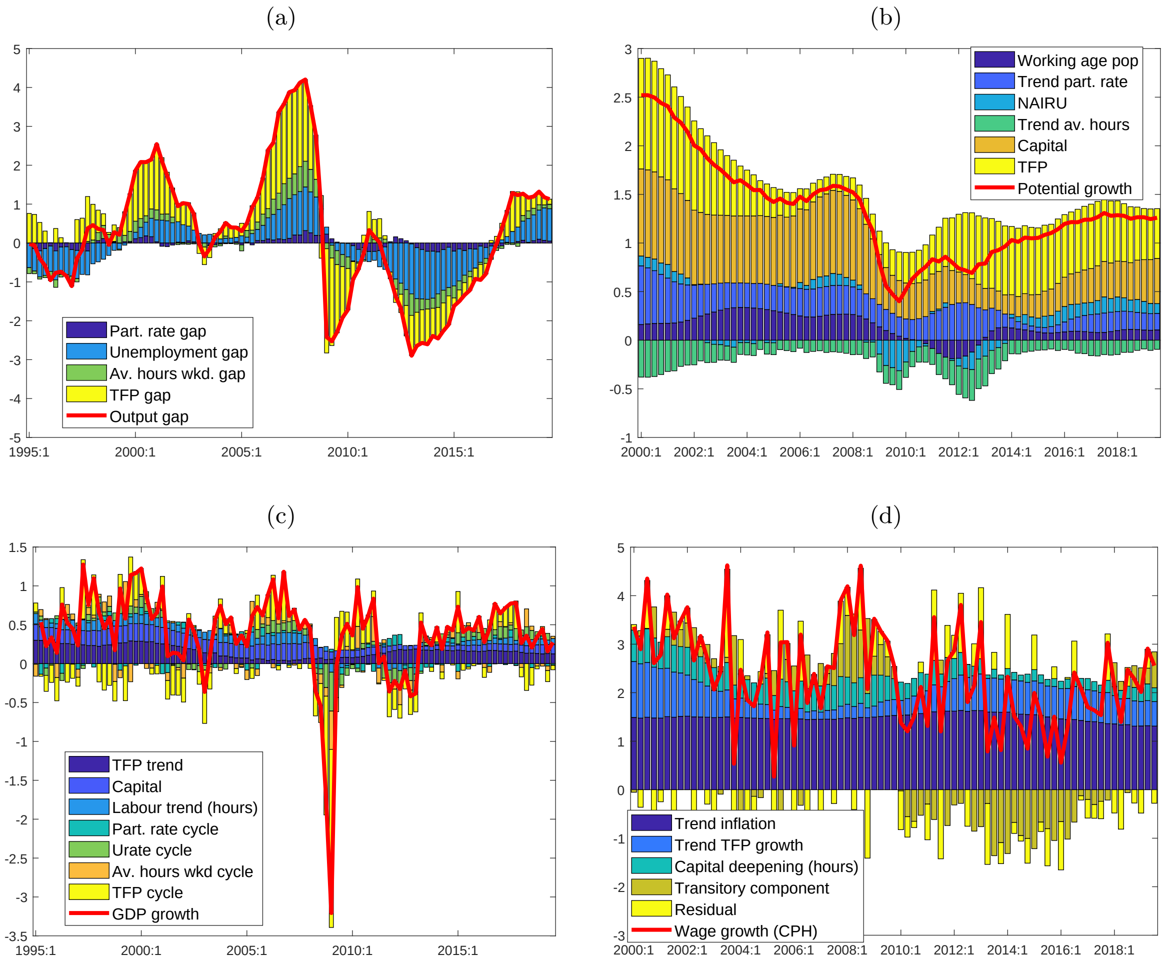


(f) Wage growth



Notes: GDP and hours worked per person are shown in natural logarithm, HP trend is a Hodrick-Prescott smoothed trend with a signal-to-noise ratio of 1/1600; unemployment rate (UNR) and the participation are reported in percentages of the labour force and working age population, respectively; core inflation (HICP excluding food and energy) and wage growth (hourly compensation) are annualised q-o-q growth rates multiplied by 100

Figure 5: Four key decompositions in Model B



Notes: the output gap is the difference between the natural logarithm of GDP and that of potential output; potential growth is shown in y-o-y growth rates; GDP growth is shown in q-o-q log differences; wage growth is the annualised q-o-q growth rate of hourly compensation. All measures are multiplied by 100

5.1 The uncertainty around state estimates

Unobservable variables can only be estimated subject to a high degree of uncertainty. This is particularly true in case of output gap estimates as noted by Orphanides and van Norden (2002). The main sources of uncertainty are the stochastic nature of the processes driving the evolution of the unobservable states (filter or state uncertainty), the precision with which parameters of the processes can be estimated (parameter uncertainty) and the possibility of model mis-specification (model uncertainty). Uncertainty is also related to revisions to past observable data.

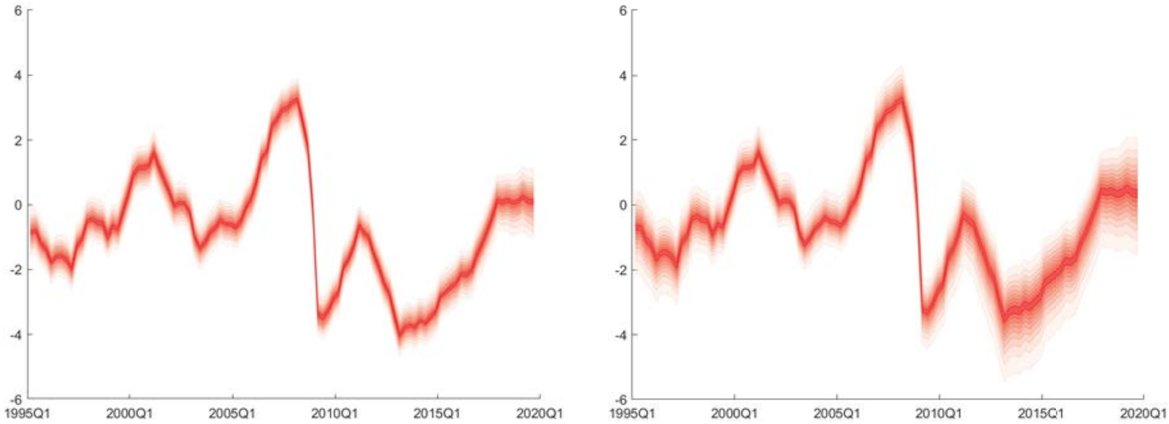
Filter uncertainty is reflected in the probabilistic nature of the state estimates that the Kalman filter (or smoother) provides. Under the assumption of linearity and Gaussian disturbances the unobserved states follow a multivariate normal distribution, where a straightforward measure of uncertainty around the point estimates (means) is the state covariance. The state covariance provided by the Kalman filter or smoother at each point in time however doesn't account for the autocovariance of the states, see e.g. Jarociński (2015). A more encompassing measure of uncertainty is provided by the simulation smoother of Durbin and Koopman (2002), which can be used to draw the shocks driving the stochastic processes of the states - conditional on the data and the parametrisation of the model - and generate the distribution of the states taking into account their covariance over time (Figure 6a).

Parameter uncertainty arises when the parameters of the underlying state-space model are unknown and have to be estimated from the data, as in most economic applications. Bayesian estimation methods provide straightforward ways for quantifying parameter uncertainty by combining prior beliefs and the information in the data to generate the posterior distribution of the parameters. In the current exercise parameter and filter uncertainty can jointly be illustrated by repeatedly drawing from the simulated posterior distribution of the parameters and employing the simulation smoother conditional on the drawn parameter vectors to draw from the distribution of the states (Figure 6b).

The possibility of mis-specification or model uncertainty is inherent in all models of the economy, which are imperfect approximations of reality. While the number of possible specifications makes it infeasible to completely hedge against model uncertainty, it is typically possible to check the most important assumptions against plausible alternatives. While model uncertainty cannot be quantified in absolute terms, Bayesian estimation makes it feasible to compare differ-

Figure 6: State and parameter uncertainty around the euro area output gap estimate (model A)

(a) Filter uncertainty (b) State and parameter uncertainty



Notes: The covariance matrix of the state shocks has been scaled by a factor which ensures the highest likelihood conditional on the coefficients in the model equations and the data. Filtered and smoothed state estimates remain unaffected by this transformation. The fanchart encompasses a 5-95 inter-percentile range in 5pp steps and was built using Deoras (2016).

ent specifications of a model that are estimated on the same data via their marginal likelihoods. Out-of-sample forecasting performance can also provide information about the severity of the mis-specification problem.

A further source of uncertainty can stem from the observable data from which the latent variables of interest are extracted. The observable data may be imprecise measures of the underlying economic concepts they represent and entail measurement errors or can be subject to revisions over time. For example, the history of national accounts data is frequently revised which can change the history of potential output estimates across data vintages. Although less frequently, but the history of Labour Force Survey based data, such as the size of the labour force, can also be subject to revisions. On the contrary, revisions to historical consumer price data tend to occur rarely, if at all. In any case, as was pointed out by Orphanides and van Norden (2002) historical data revisions are not the most important sources of revisions to output gap estimates. Furthermore, changes in estimates of unobserved states in response to historical data revisions can be even considered as desirable, as the revised data reflects more accurately the sate of the economy and the resulting revisions are unrelated to the properties of the underlying model[12]. Data revisions can be an issue however, if they are sensitive to the business cycle.

[12]The estimation sample is often expanded by off-model forecast data that are treated as observables. Since forecast revisions are typically larger than revisions to historical data and involve several periods beyond the end of the historical data sample their impact on smoothed state estimates can be larger than that of revisions to past data

The key concern of policymakers related to the uncertainty surrounding policy relevant unobservables is their tendency to be revised over time when new observations are added to the estimation sample - even in the absence of historical data revisions, as pointed out by Orphanides and van Norden (2002). By definition, these revisions occur only in the case of smoothed state estimates, since these are estimated on the basis of the full sample information. The fixed interval smoothers typically applied in the economics literature mechanically update past state estimates when new observations are added to the sample. From the Kalman filtering/smoothing algorithms it follows that the magnitude of the backward update of state estimates depend on the prediction error of the model with respect to the new observations, see e.g. Harvey (1990). Thus, in the absence of historical data revisions, the revision properties of estimates of unobservable states depend on the forecasting performance of the underlying model in terms of the observables that have the largest weights implicitly assigned to the states, in the sense of Koopman and Harvey (2003). To asses how these uncertainties affect state estimates - in particular the output gap - over time the following section conducts a thorough evaluation exercise.

6 An evaluation exercise

The evaluation exercise introduced in this chapter focuses on the output gap, which is the most widely used measure of slack in the economy. The usefulness of an output gap measure for policy purposes depends on a number of factors as pointed out, e.g., by Camba-Mendez and Rodriguez-Palenzuela (2003). First, the output gap measure should tell a plausible story about the state of the economy in the country or region for which it was estimated. Second, it should not be subject to excessive revisions in response to new data and should paint a consistent picture of the business cycle over time. Third, in a central banking context, it is important that the output gap measure has a proven ability for predicting inflation. Other possible criteria, such as the symmetry of output gap measures or a zero mean requirement are also relevant, but hard to assess if the sample size is relatively low and the estimates cannot possibly include more than a couple of full cycles.

Since the plausibility of a potential output or output gap measure is a qualitative category, it cannot be measured directly. However, plausibility is typically assessed against some implicit theory or assumptions implying co-movements across the estimated output gap and some other, possibly observable, indicators. Often, the assessment of plausibility is based on an assumed link between the estimated output gap measure and the nominal side of the economy, labour market developments or some observable cyclical indicators as in the case of the European Commission's plausibility tool, see Hristov et al. (2017). Of course, these cyclical indicators are necessarily imperfect measures of the business cycle, otherwise it would be pointless to use more complicated algorithms to estimate the latter. Ideally, to the extent these indicators contain useful information about the business cycle, they should be embedded in the model that produces the output gap estimate.

The revision properties of an output gap measure over time cannot be precisely assessed in quantitative terms, since this would imply the existence of an objective benchmark, i.e. an observable or 'true' measure of the output gap, rendering the efforts to estimate it redundant. In the absence of a 'true' output gap measure a working assumption could be that the latest vintage from a set of output gap estimates over an expanding sample is the best guess for a benchmark to assess the magnitude of past revisions. The assessment of the inflation forecasting performance of an output gap measure can be done in a number of different ways. One can use the vintages of the output gap (or other slack measures) in standalone, reduced form Phillips curves and assess their forecasting performance conditional on the realisations and projections of other explanatory variables, as done by Nickel et al. (2019). Another possibility is to use the underlying model from which the output gap estimate stems to create out-of-sample forecasts of inflation – provided that the model includes a measure of inflation as an observable and features a meaningful relationship between inflation and the output gap, as e.g. in Jarociński and Lenza (2018) or the current exercise.

Since modelling choices are made with the benefit of the hindsight and the modellers' prior beliefs cannot be fully detached from past experience, it is not possible to re-create exactly the same information sets that prevailed in past periods in case of more complicated models. Keeping this in mind, it is desirable to approximate a real-time context as much as possible when assessing the the revision properties and/or forecasting performance of a model. A necessary condition for such a real-time exercise is that the extraction of the trend and cyclical components in each period is done conditional on information available up to that period. Barring historical data revisions, this is equivalent to running a one-sided filter on the latest available data sample. It is then possible to construct a measure of revision properties by comparing one-sided and two-sided estimates. A true real time context can further be approximated by calculating the one-sided estimates by re-estimating the parameters of the underlying model in each time period.

To this end in the following exercise smoothing and the estimation of model parameters are done on an expanding sample, starting with an initial sample of sufficient length that allows the estimates of the unobservables to stabilise. Since the revision and forecasting properties of the model are compared to alternatives that can be recursively implemented over the same expanding sample, there is no need to use real-time data vintages to assess relative performance. Building the exercise on real-time data vintages would be important if the revisions or forecasting performance were compared to real time estimates published in the periods in focus.

In the following the revision properties and forecasting performance of the model specifications introduced in this paper are assessed based on the quantitative criteria outlined above. In particular the revision properties in terms of the root mean squared error (RMSE) of the estimated output gap measures with respect to the latest, full sample estimate are compared to univariate benchmarks. The (unconditional) forecasting performance of the model with respect to its observable input variables is also assessed in terms of root mean squared forecast errors (RMSFE) in comparison with a set of simple benchmarks.

6.1 Revisions of the output gap in response to new data points

This part assesses the revision properties of the output gap estimates of models A and B against a set of univariate benchmarks, such as the Hodrick-Prescott filter a band-pass filter of the type of Christiano and Fitzgerald (2003), and a local level model as described in Chapter 2.2 of Harvey (2006). The results are based on an expanding window exercise, with (a vector of) new observations added in each time period in a pseudo real-time context (i.e. the data vintage remains the same in each period). The starting point of the subsample windows over which the evaluation is conducted is always Q1 1995. The end points span from Q1 2000 to Q3 2019 in quarterly steps. The first window end point was selected to allow a sufficiently long sample for the dissipation of the effect of the initial conditions on state and parameter estimates on the one hand, and the assessment of the revision properties before and during the great financial crisis on the other. To reflect on more recent developments, Appendix D assesses the output gap revisions in response to the Covid-19 shock in 2020.

The evaluation exercise is based on two sets of results that are generated from model A and B. In the first set, the parametrisation of the models are kept unchanged at the values estimated over the most recent sample (ending in Q3 2019). In the second set model parameters are re-estimated in each period and evaluated at the median of their posterior distributions.

The difference between the sample average of headline HICP inflation and core inflation (HICP excluding food and energy), which plays a role in the price and wage blocks of the model A, is also recalculated for each subsample. Table 5 shows the RMSE and the mean deviation (i.e. bias) of the output gap estimates *vis-a-vis* their full sample estimates. The RMSEs in the first row of the table are calculated over the deviations of the estimated endpoints of the output gap in each subsample compared to the final, full sample (smoothed) estimate in the same time period. In the second row the RMSEs are based on deviations over the entire sample, i.e. they take into account the history of the output gap series over each subsample windows. Figure 7 shows the full output gap trajectories.

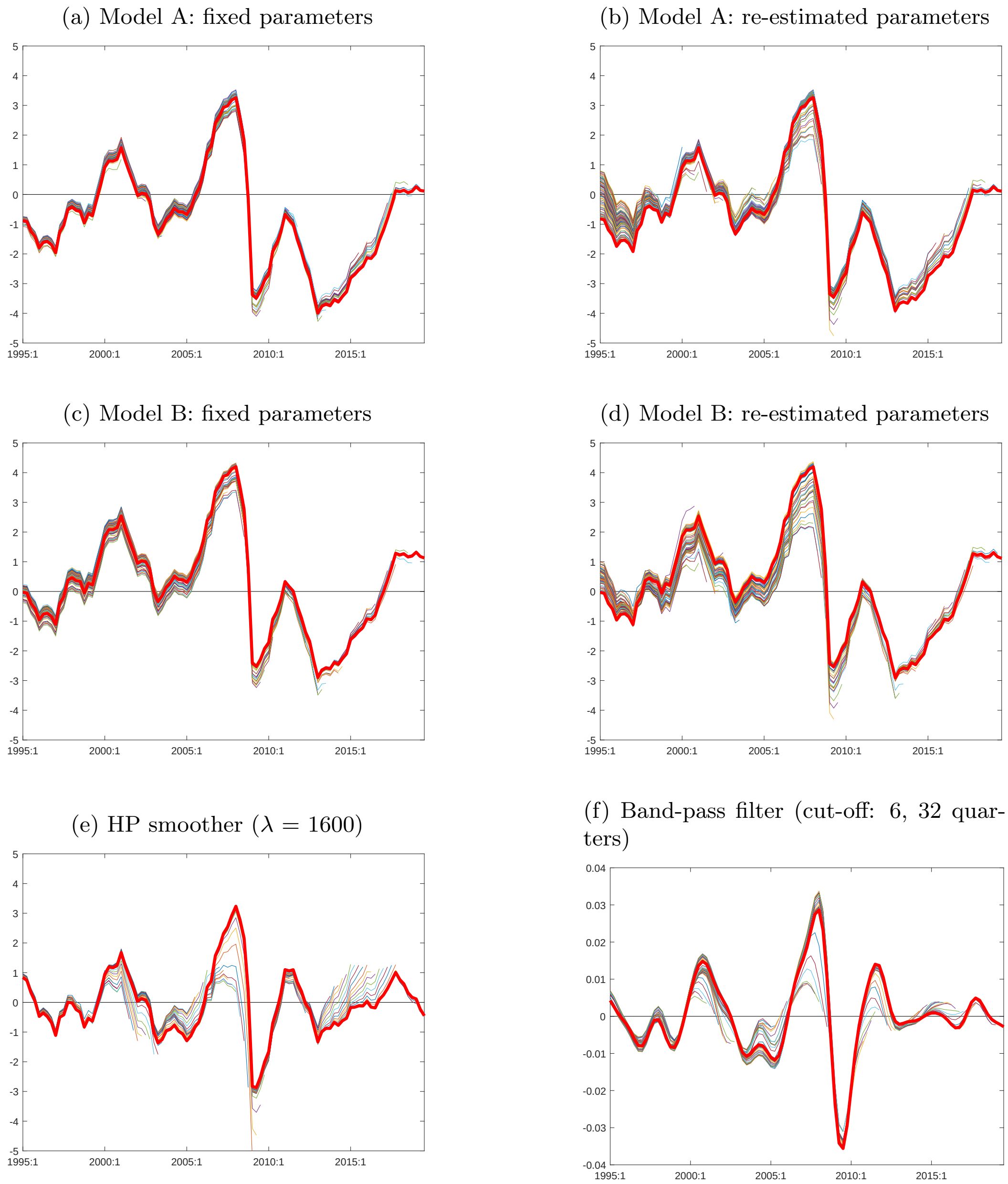
Table 5: Root mean squared error and bias of output gap estimates

	A1	A2	B1	B2	HP	BP	LL
RMSE total	0.166	0.268	0.212	0.452	0.344	0.259	0.581
RMSE endpoints	0.285	0.576	0.359	0.846	1.170	0.764	1.981
Bias total	0.067	0.074	-0.045	-0.246	-0.015	-0.015	0.262
Bias endpoints	0.048	0.017	-0.157	-0.507	-0.130	-0.242	1.707

Note: Log points (log deviations multiplied by 100). A1: model A with parameters kept constant over time. A2: parameters re-estimated in each period and evaluated at their posterior median. B1-B2 refer to similar setups in case of model B. HP refers to the Hodrick-Prescott filter, with signal-to-noise ratio of 1/1600, BP to a Christiano-Fitzgerald type band-pass filter with cut-off periodicities of 6 and 32 quarters and LL to a local level filter with a signal-to-noise ratio of 1/1600.

Overall, the fixed parameter output gap estimates from model A and B (columns A1 and B1 in Table 5) outperform the benchmarks in terms of RMSE, both in case of total (i.e. with prediction errors calculated over each quarter in the expanding window horizon) and endpoint revisions. In case of model A this remains true even when taking into account parameter uncertainty (column A2), except for total RMSE against the band-pass filter. In terms of bias the picture is less clear-cut. In terms of total deviations, model A tends to slightly overestimate the final output gap in real-time, on average by around 0.07 log points, both with fixed and re-estimated parameters, while model B tends to have a stronger negative bias, particularly with re-estimated parameters. The HP and band-pass filtered benchmarks underestimate the final output gap, but have the smallest bias in absolute terms. Regarding the magnitude of the bias with respect to the endpoints, model A clearly outperforms model B and all the benchmarks.

Figure 7: Output gap revision properties



(a) Model A: fixed parameters

(b) Model A: re-estimated parameters

(c) Model B: fixed parameters

(d) Model B: re-estimated parameters

(e) HP smoother ($\lambda = 1600$)

(f) Band-pass filter (cut-off: 6, 32 quarters)

Note: Expanding window smoothed estimates, first window: 1995Q1-2000Q1, last window 1995Q1-2019Q3. "Fixed parameters" refer to a parameter set estimated over the full sample and kept fixed, "re-estimated" refers to model parameters re-estimated and evaluated at their posterior medians in each quarter between 2000Q1 and 2019Q3. The bold red line indicates to the full sample estimate, the thin lines show estimates over the expanding windows

6.2 Forecasting performance

This subsection examines the out of sample forecasting performance of model A and model B over against a set of benchmarks. While the models introduced in this paper were not built with forecasting as the main purpose, forecasting performance is an important criterion against which models aiming at estimating unobservable variables can be evaluated quantitatively. Furthermore, as noted earlier, from a central banking perspective, the inflation forecasting ability of slack measures is of key importance, particularly over medium term horizons. However, the overall goodness of fit of models can be assessed on the basis of their general predictive ability, that is, with respect to all their observable variables. The evaluation exercise is conducted over multi-step forecast horizons spanning form one to eight quarters, and is based on iterated one-step ahead forecasts. Otherwise the set-up of the exercise is similar to that of the revision evaluation exercise: pseudo out-of-sample forecasts are generated in an expanding window context, on a single data vintage, with Q2 2000 as the first projected quarter and Q3 2019 as the last one. The forecast horizons are tapered off at the end of the sample, so that the last prediction error taken into account in the exercise is based on the one quarter ahead forecast for the sample end point. Forecasting performance is evaluated using RMSFEs calculated over the pool of prediction errors for all horizons as well as for each horizon separately.

The forecast evaluation exercise employs the following set of benchmarks: an ARIMA model fitted to each observable variables, with the specification (i.e. the number of AR and MA terms and the degree of integration) selected on the basis of the full sample. First, unit root tests are run on the level and first difference of each individual series, in particular the augmented Dicky-Fuller test and the KPSS the test. A degree of integration is selected if both tests agree about it. The number of AR and MA terms are selected on the basis of the Bayesian info criterion, assuming that neither is larger than four. The resulting ARIMA specifications are kept fixed over the subsamples. The second benchmark is an unrestricted VAR(1) estimated on GDP, the unemployment rate, core inflation and wage growth. These four variables were selected due their role in the three key relationships of model A and B, i.e. the price and wage Phillips curves as well as Okun's law. For the purpose of estimating the VAR, the variables are transformed according to the degree of integration found by the unit root tests in the ARIMA case. In particular, GDP is first-differenced, while the other three variables are kept in levels. The third benchmark is based on the state space representation of the HP filter, with the usual smoothing parameter for quarterly data (1600)[13], over which the Kalman filter is applied to produce the one to eight quarters ahead iterated forecasts.

[13]In particular, following Harvey and Trimbur (2008) a local linear trend plus noise model is applied, with the standard deviation of level shocks set to zero and the relative variance of the trend shock and the noise is set to 1/1600.

Table 6 summarises the forecasting performance of model A and B as well as three benchmarks in the case of model parameters remaining fixed at their full sample estimates over the expanding windows. Model B and model A have an advantage over the benchmarks in terms of forecasting the level of GDP. In terms of inflation the VAR and model A have the lowest RMSFE. The VAR performs best in forecasting wage growth, while the ARIMA benchmark has the best performance in forecasting the unemployment rate. In terms of the remaining three observables, model B performs best in case of average hours worked, while the ARIMA benchmark has the lowest RMSFE in case of the labour force participation rate, the capital stock and working age population. The HP filter consistently underperforms against the alternative approaches.

Table 6: 1 to 8 quarters ahead RMSFEs, fixed parameters

	mod. A	mod. B	VAR(1)	ARIMA	HP
y	1.683	1.635	1.790	1.934	2.824
π	0.127	0.141	0.125	0.133	0.179
w	0.237	0.232	0.217	0.235	0.304
u	0.868	0.877	0.776	0.667	1.435
ah	0.518	0.491	-	0.518	0.723
lp	0.482	0.476	-	0.323	0.477
k	0.309	0.309	-	0.251	0.847
wp	0.157	0.157	-	0.129	0.388

Note: Log points (log deviations multiplied by 100) and percentage points. Fixed parameters refer to full sample estimates kept fixed over the expanding windows. HP refers to forecasts produced by the Hodrick-Prescott filter, with a signal to noise ratio fixed at 1/1600.

Table 7 shows the predictive ability of model A and B and the benchmark models with parameters re-estimated in each period of the expanding window (since the signal-to-noise ratio of the HP filter is not estimated, it is excluded from this table). The relative performance of the models remain broadly similar to the fixed parameter case. Model A and B retain their advantage in predictive ability in terms of GDP vis-a-vis the benchmarks and model A has now the best inflation forecasting performance, although by a negligible margin. In cases where model A or B are not the best performers, their performance gaps in terms of RMSFE became narrower.

Regarding the horizon-specific performance (shown in Appendix C), model A and B tend to produce larger RMSFEs over the shorter horizons, but their relative performance improves as the length of the horizon increases. In terms of forecasting GDP, model B is the best performer

Table 7: 1 to 8 quarters ahead RMSFEs, re-estimated parameters

	mod. A	mod. B	VAR	ARIMA
y	1.898	1.840	3.126	2.172
π	0.152	0.187	0.176	0.153
w	0.252	0.256	0.269	0.242
u	0.920	0.960	1.121	0.886
ah	0.799	0.844	-	0.542
lp	0.507	0.500	-	0.434
k	0.309	0.309	-	0.320
wp	0.157	0.157	-	0.188

Note: Log points (log deviations multiplied by 100) and percentage points. Parameters re-estimated in each expanding window.

over all horizons, closely followed by model A, both with fixed and re-estimated parameters. In terms of forecasting inflation, model A is outperformed by the ARIMA and VAR benchmarks only at the one quarter ahead horizon in the fixed parameter case. In the re-estimated case it has the advantage at horizons beyond 4 quarters. The pattern of improving relative forecasting performance as the horizon grows holds in general in the case of the remaining variables as well.

Since the only difference between model A and B is the formulation of the trend inflation process, their relative performance with respect to forecasting inflation has important implications about this modelling choice. Over the full evaluation sample model A has the superior inflation forecasting performance, consistently over each projection horizon (see Appendix C). However, model B gains the upper hand when the evaluation exercise is constrained to a post-2010 sample (Table 8), when headline inflation was well below 2 per cent for long periods. Furthermore, the change in relative performance is related to the longer forecast horizons. This suggests that model B - featuring a stochastic trend inflation process - can better accommodate the observed decline in (core) inflation, despite the fact that it yields a less negative output gap on average over the same time interval. Since model A and B have been estimated using Bayesian methods over the same data set, their fit can be also be assessed via comparing their marginal likelihoods (or marginal data densities), which is a metric based on one-step ahead prediction errors. Assuming identical prior odds, the modified harmonic mean estimator of the log marginal likelihoods, see Geweke (1999), implies a Bayes factor of around 6.3 in favour of model B.

Table 8: Model comparison: inflation trend specifications, post-2010 sample

	relative RMSFE							
horizon (quarters)	1	2	3	4	5	6	7	8
Model A/B fixed	0.92***	0.95	0.95	1.13	1.13	1.23*	1.36**	1.45**
Model A/B re-est.	0.91***	0.89	0.87	0.97	1.02	1.11	1.32	1.50**

Note: ***.** and * refer to 1%, 5% and 10% significance levels of rejecting the null hypothesis of equal predictive ability under the Diebold-Mariano test

7 Concluding remarks

This paper built a multivariate unobserved components model based on a Cobb-Douglas production function and a set of economic relationships, such price and wage Phillips curves and a version of Okun's law to estimate key unobservable variables of policy relevance. The model features labour market hysteresis effects and a relationship between trend wage growth, trend inflation and trend productivity growth as well. Two specifications of the model were estimated on aggregate euro area data, differing in terms of stochastic process governing the evolution of trend inflation. The first, baseline specification had a fixed anchor, consistent with the ECB's price stability objective (model A), the second one assumed a stochastic trend in inflation (model B).

The models were estimated via Bayesian methods and the estimates of the unobserved states were provided by a Kalman smoother algorithm. The data seems to contain information about most of the model parameters, in particular the data supports the prior belief in the existence of the Phillips curves, Okun's law and labour market hysteresis effects. The resulting output gap estimates suggest that the trend inflation assumption is a crucial one. The anchored trend inflation assumption yields a more negative output gap after the great financial crisis, which became broadly neutral in the first half of 2018 and remained so until the second half of 2019. The stochastic trend specification suggest a decline in the trend component of inflation since 2013 and is consistent with a less negative output gap after the financial crisis and a positive one since the second half of 2018.

The output gap measures have good revision properties against a set of benchmarks, such as the Hodrick-Prescott filter or a band-pass filter, in an expanding window setup over a single data vintage. Furthermore, both model specifications have reasonable forecasting performance in terms of the observables entering the model against a VAR and ARIMA benchmarks. They perform particularly well in terms of forecasting GDP growth and core inflation over 1 to 8 quarters horizons. While model A performs better in terms of output gap revisions and inflation forecasting over the full sample, model B gains the upper hand in forecasting performance after 2011 Q1. There is also some moderate evidence in favour of model B on according to marginal likelihoods. This suggest that a decline in trend inflation may be a reasonable explanation for the low-inflation environment prevailing in the second half of the 2010s, in spite of disappearing slack. Based on the evaluation exercise, the production function based multivariate unobserved components model developed in this paper can be a useful tool in a monetary policy context.

A State Space representation of the baseline model

$$
\begin{pmatrix} y_t \\ u_t \\ k_t \\ wp_t \\ ah_t \\ lp_t \\ \pi_t \\ w_t \end{pmatrix}
=
\begin{pmatrix}
1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1
\end{pmatrix}
*
\begin{pmatrix} \hat{y}_t \\ \hat{y}_{t-1} \\ \bar{y}_t \\ \widetilde{a}_t \\ \hat{u}_t \\ \hat{u}_{t-1} \\ \bar{u}_t \\ \tilde{u}_t \\ \hat{lp}_t \\ \bar{lp}_t \\ \tilde{lp}_t \\ \hat{ah}_t \\ \bar{ah}_t \\ \tilde{ah}_t \\ \bar{wp}_t \\ \tilde{wp}_t \\ \bar{k}_t \\ \tilde{k}_t \\ \hat{\pi}_t \\ \bar{\pi}_t \\ \hat{w}_t \\ \bar{w}_t \end{pmatrix}
$$

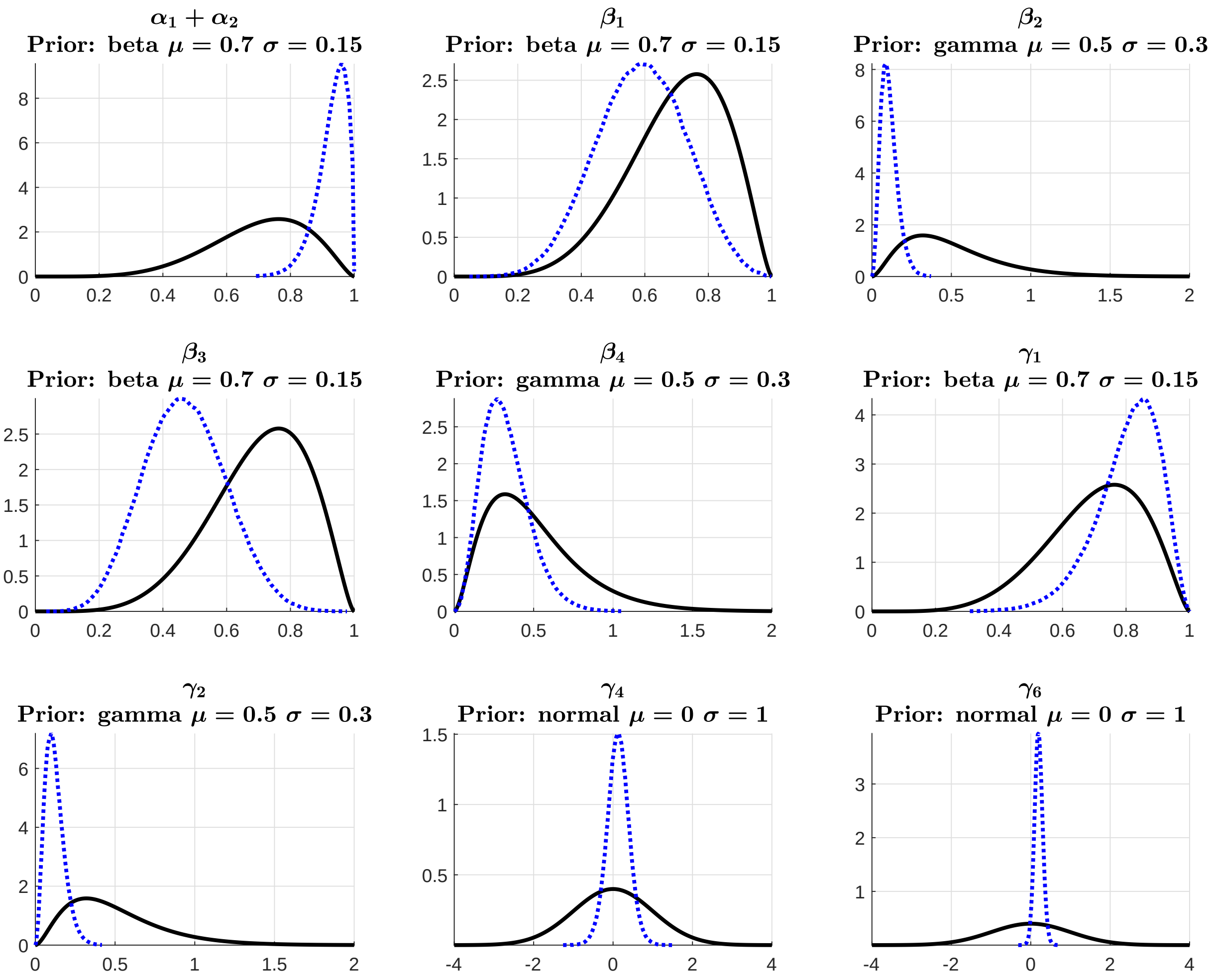
$$
\begin{pmatrix} \hat{y}_t \\ \hat{y}_{t-1} \\ \bar{y}_t \\ \widetilde{a}_t \\ \hat{u}_t \\ \hat{u}_{t-1} \\ \bar{u}_t \\ \tilde{u}_t \\ \hat{lp}_t \\ \bar{lp}_t \\ \tilde{lp}_t \\ \hat{ah}_t \\ \bar{ah}_t \\ \tilde{ah}_t \\ \bar{wp}_t \\ \tilde{wp}_t \\ \bar{k}_t \\ \tilde{k}_t \\ \hat{\pi}_t \\ \bar{\pi}_t \\ \hat{w}_t \\ \bar{w}_t \end{pmatrix}
=
\begin{pmatrix}
1+\alpha_1 & -\alpha_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & -\nu\kappa_1 & \nu\kappa_1 & 0 & 0 & 0 & 0 & \nu(1-\kappa_2) & 0 & 0 & 0 & 0 & \nu & 0 & 1-\nu & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\gamma_2 & 0 & 0 & 0 & \gamma_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \kappa_1 & -\kappa_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \kappa_1 & -\kappa_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\gamma_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\gamma_6(1+\alpha_1) & \gamma_6(-\alpha_2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\kappa_2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\kappa_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\beta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi & 0 & 0 \\
0 & 0 & 0 & 0 & -\beta_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_3 & 0 \\
0 & 0 & 0 & 1 & (1-\nu)\kappa_1 & (\nu-1)\kappa_1 & 1-\nu & 0 & 0 & \nu-1 & (\nu-1)(1-\kappa_2) & 0 & (\nu-1) & 0 & (\nu-1) & (\nu-1) & 0 & (1-\nu) & 0 & \phi & 0 & 0
\end{pmatrix}
\begin{pmatrix} \hat{y}_{t-1} \\ \hat{y}_{t-2} \\ \bar{y}_{t-1} \\ t\tilde{f}p_{t-1} \\ \hat{u}_{t-1} \\ \hat{u}_{t-2} \\ \bar{u}_{t-1} \\ \tilde{u}_{t-1} \\ \hat{lp}_{t-1} \\ \bar{lp}_{t-1} \\ \tilde{lp}_{t-1} \\ \hat{ah}_{t-1} \\ \bar{ah}_{t-1} \\ \tilde{ah}_{t-1} \\ \bar{wp}_{t-1} \\ \tilde{wp}_{t-1} \\ \bar{k}_{t-1} \\ \tilde{k}_{t-1} \\ \hat{\pi}_{t-1} \\ \bar{\pi}_{t-1} \\ \hat{w}_{t-1} \\ \bar{w}_{t-1} \end{pmatrix}
+
$$

$$
+\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ (1-\phi)\pi^* \\ 0 \\ (1-\phi)\pi^* + wd \end{pmatrix}
+\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & \nu & 0 & \nu & 0 & \nu & \nu & (1-\nu) & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\gamma_6 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1-\nu & 0 & \nu-1 & 0 & \nu-1 & \nu-1 & 1-\nu & 0 & \phi & 0 & 1
\end{pmatrix}
\begin{pmatrix} \varepsilon_{\hat{y},t} \\ \varepsilon_{\tilde{a},t} \\ \varepsilon_{\hat{u},t} \\ \varepsilon_{\tilde{u},t} \\ \varepsilon_{\hat{lp},t} \\ \varepsilon_{\tilde{lp},t} \\ \varepsilon_{\hat{ah},t} \\ \varepsilon_{\tilde{ah},t} \\ \varepsilon_{\tilde{wp},t} \\ \varepsilon_{\tilde{k},t} \\ \varepsilon_{\hat{\pi},t} \\ \varepsilon_{\tilde{\pi},t} \\ \varepsilon_{\hat{w},t} \\ \varepsilon_{\tilde{w},t} \end{pmatrix}
$$

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon) \qquad \boldsymbol{\alpha_0} = \begin{pmatrix} 0 \\ 0 \\ \bar{y}_0 \\ \tilde{a}_0 \\ 0 \\ 0 \\ \bar{u}_0 \\ \tilde{u}_0 \\ 0 \\ \bar{lp}_0 \\ \tilde{lp}_0 \\ 0 \\ \bar{ah}_0 \\ \tilde{ah}_0 \\ \bar{wp}_0 \\ \tilde{wp}_0 \\ \bar{k}_0 \\ \tilde{k}_0 \\ 0 \\ \bar{\pi}_0 \\ 0 \\ \bar{w}_0 \end{pmatrix} \qquad \boldsymbol{P_0} = \begin{pmatrix} 0.0025 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0.0025 \end{pmatrix}$$

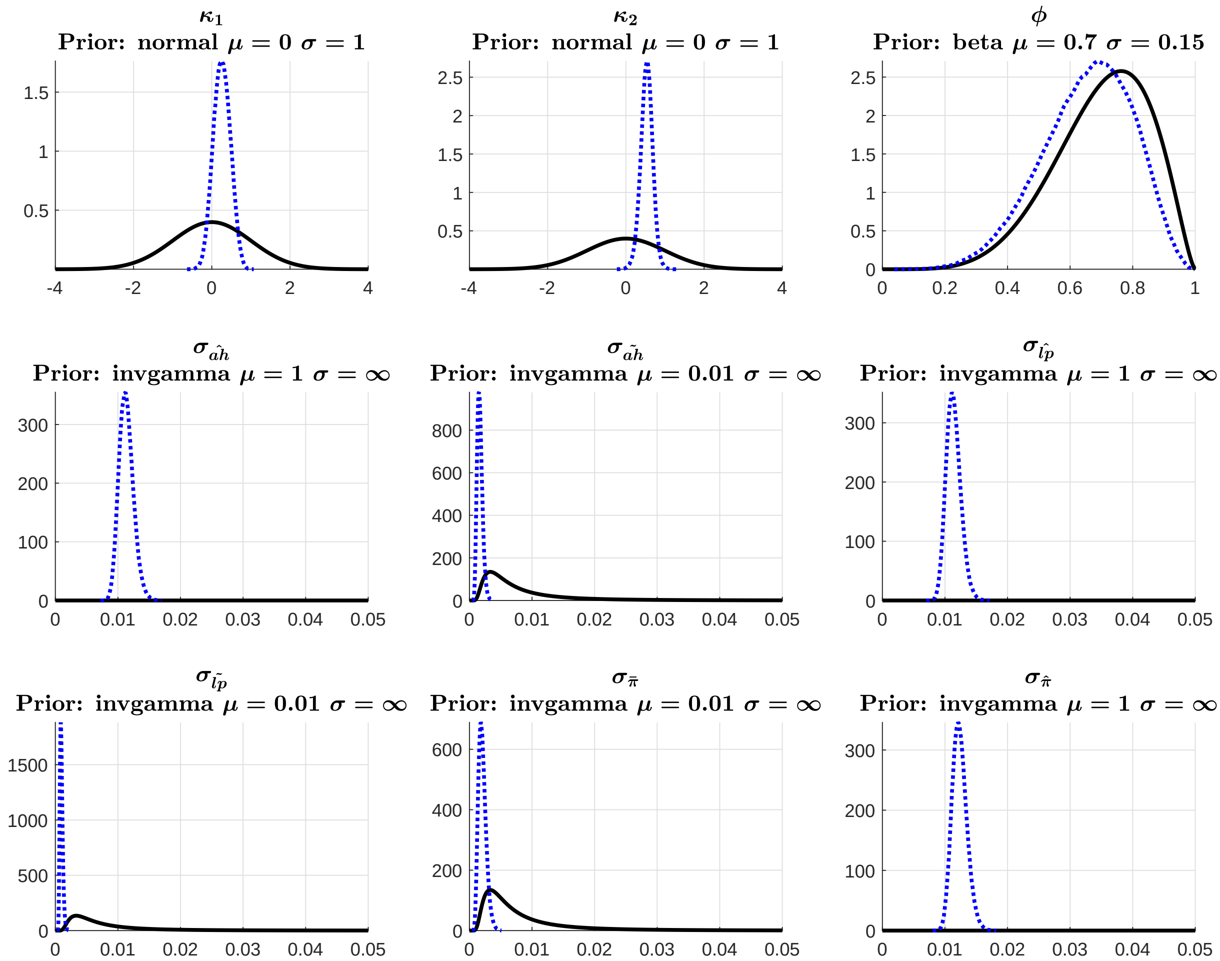
B Posterior statistics

Figure B1: Prior and marginal posterior distribution of the parameters (model A)



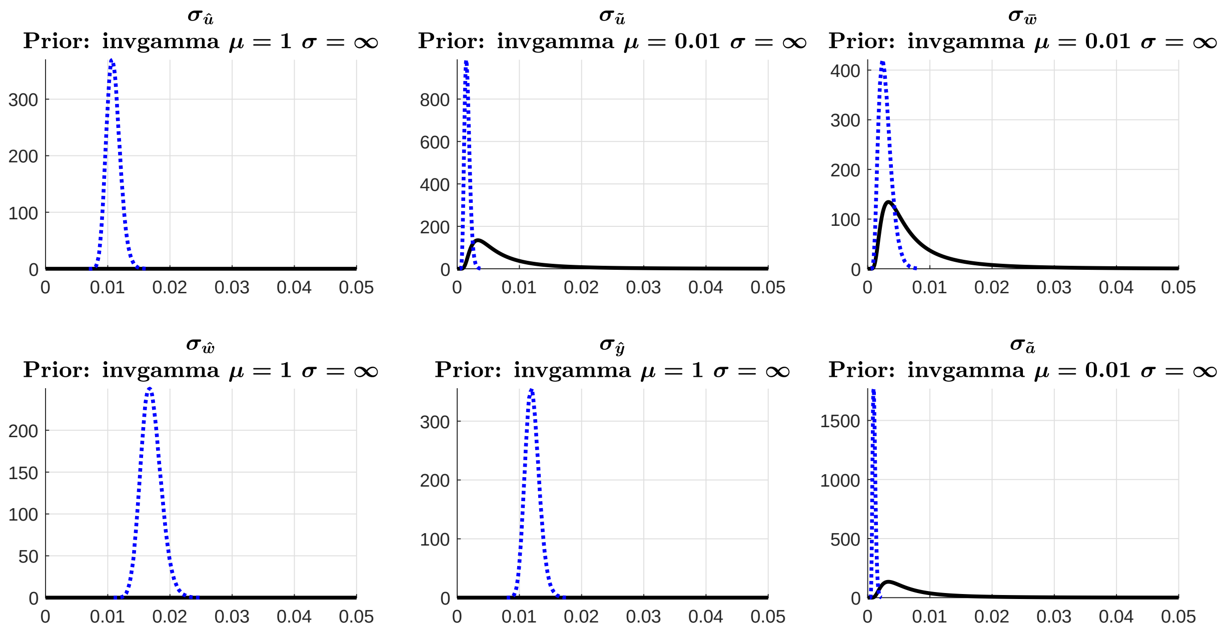
Notes: black solid line: prior distribution, blue dotted line: posterior distribution

Figure B2: Prior and marginal posterior distribution of the parameters (model A)



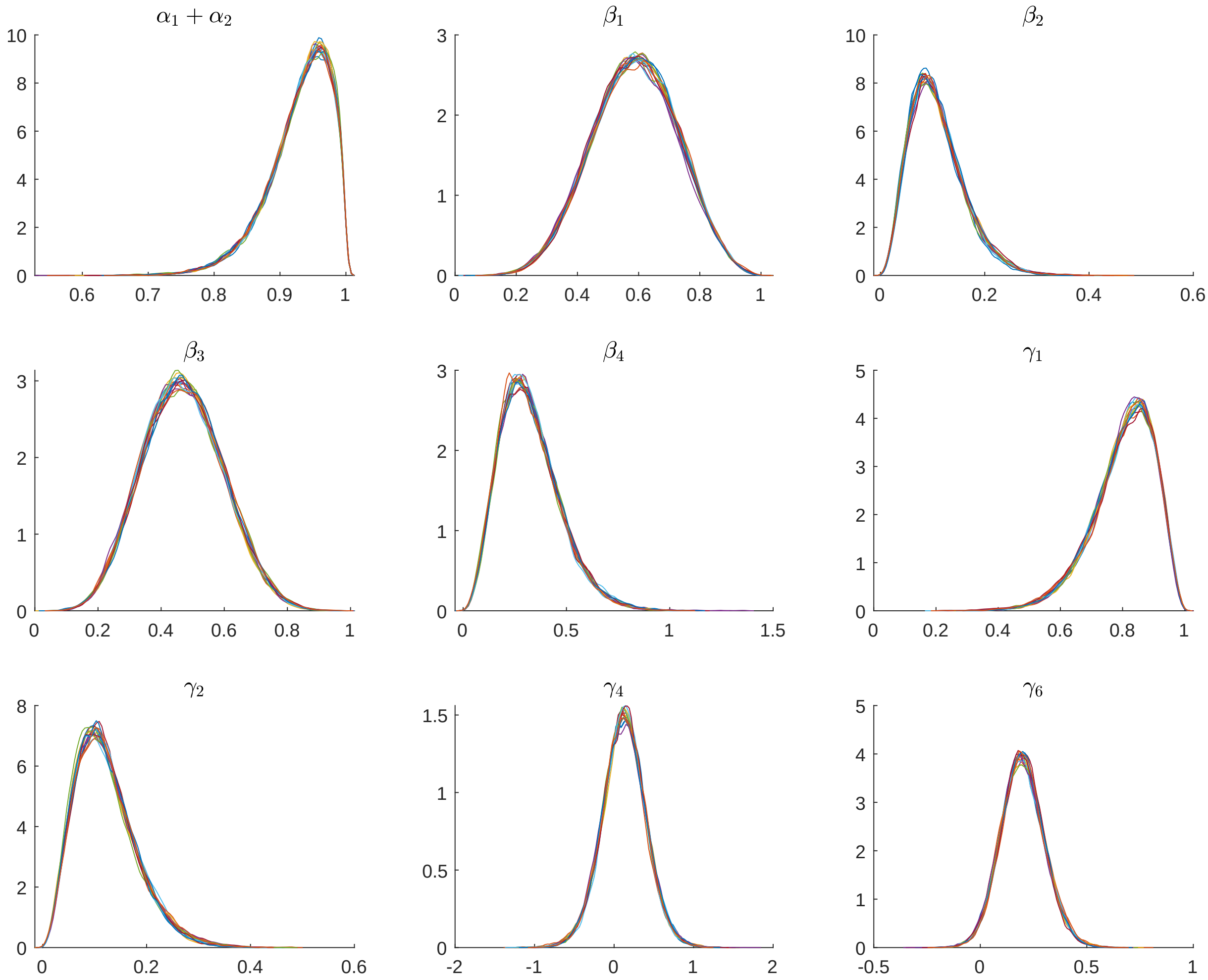
Notes: black solid line: prior distribution, blue dotted line: posterior distribution

Figure B3: Prior and marginal posterior distribution of the parameters (model A)



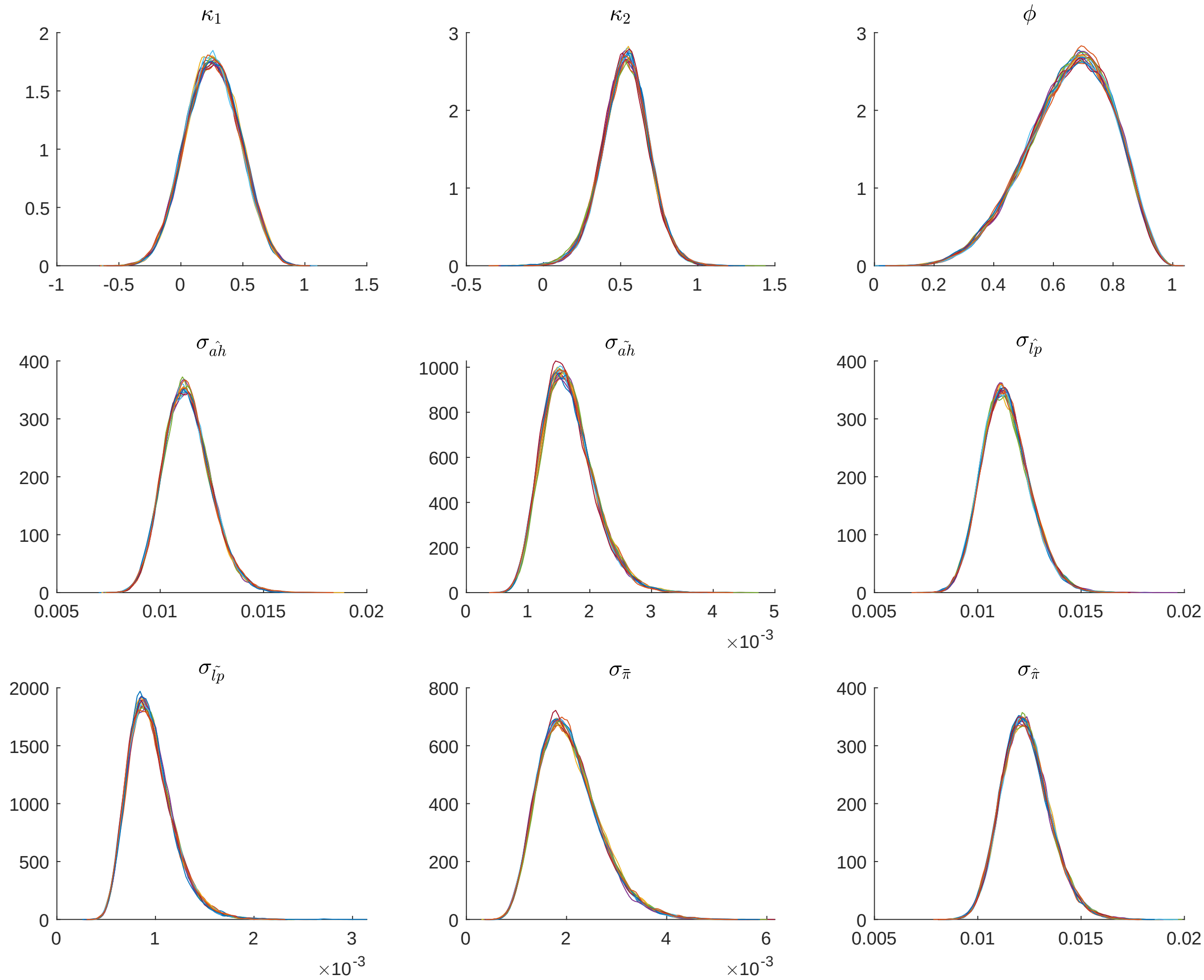
Notes: black solid line: prior distribution, blue dotted line: posterior distribution

Figure B4: Posterior distributions - parallel posterior chains (model A)



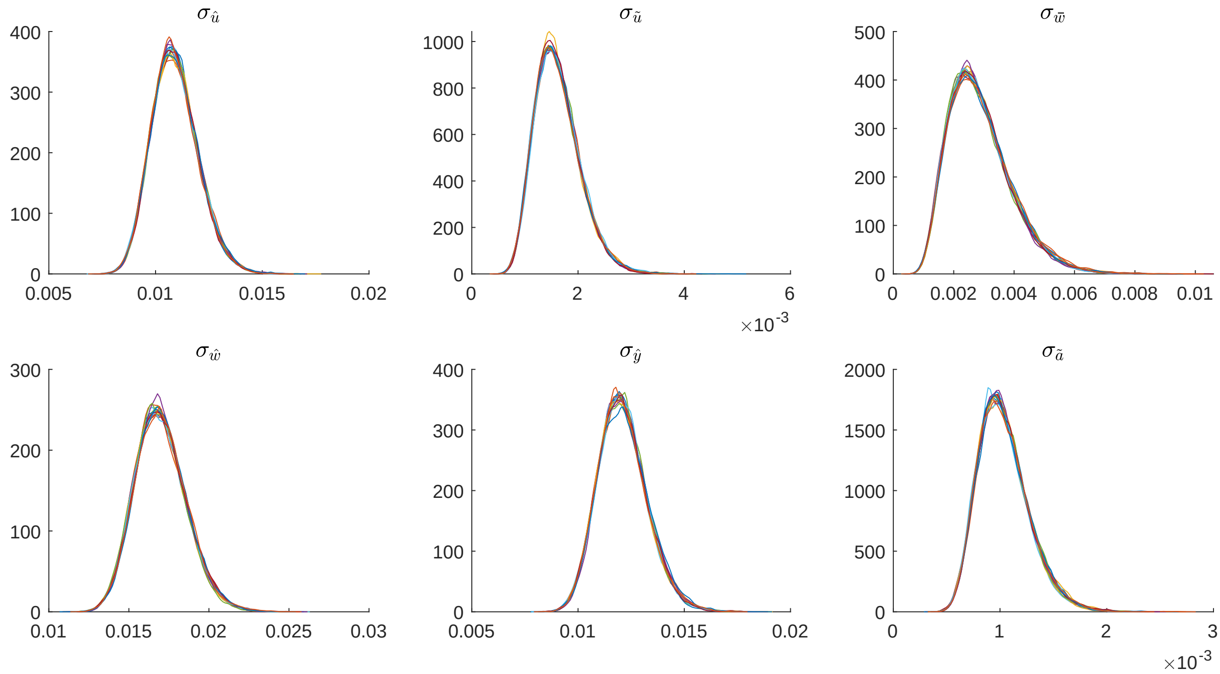
Notes: 16 parallel simulated posterior chains for each parameter

Figure B5: Posterior distributions - parallel posterior chains (model A)



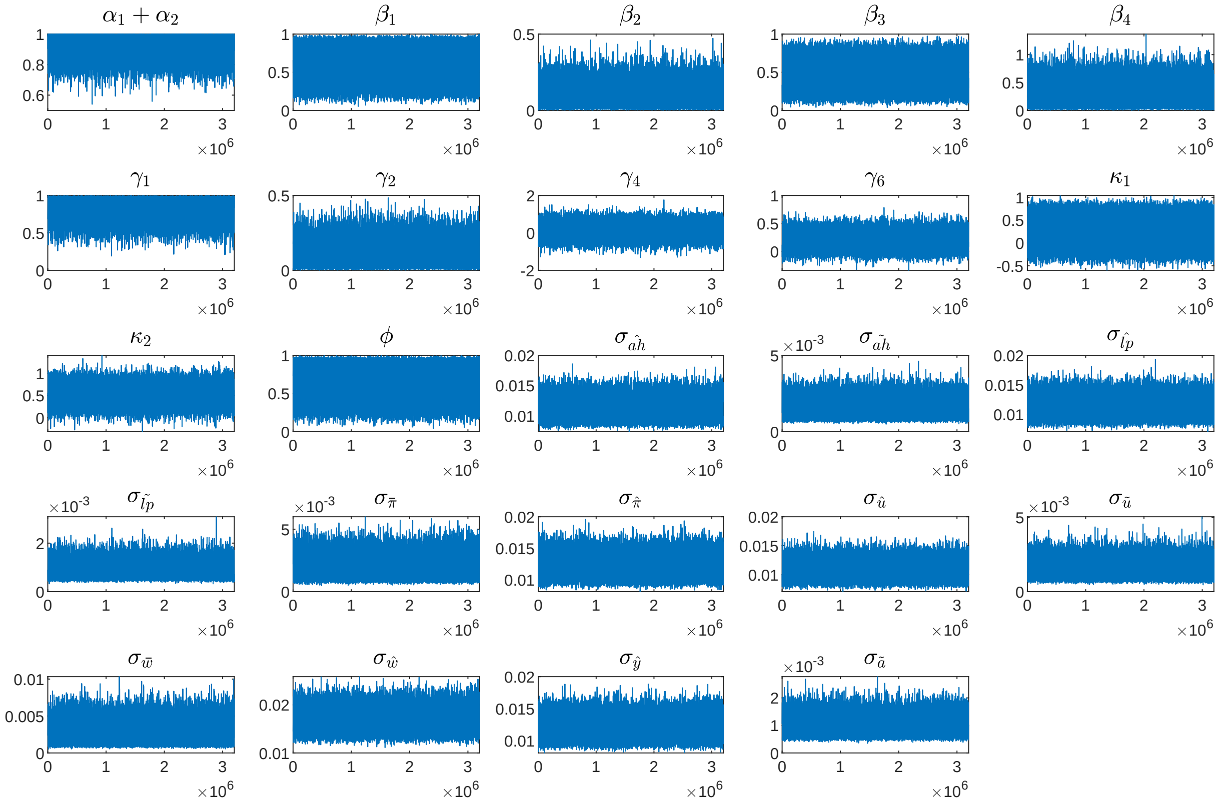
Notes: 16 parallel simulated posterior chains for each parameter

Figure B6: Posterior distributions - parallel posterior chains (model A)



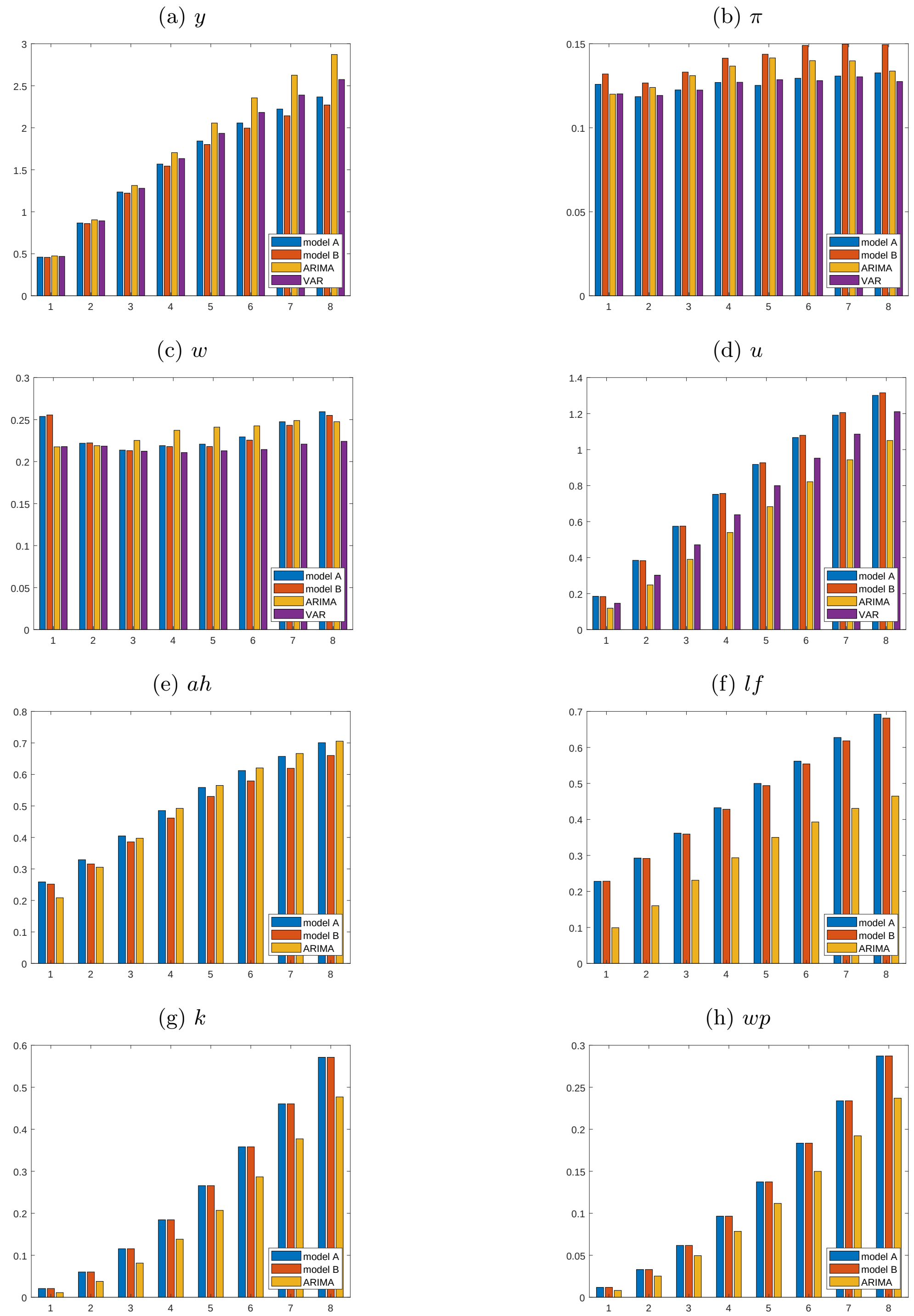
Notes: 16 parallel simulated posterior chains for each parameter

Figure B7: Trace plots of simulated posterior chains (model A)



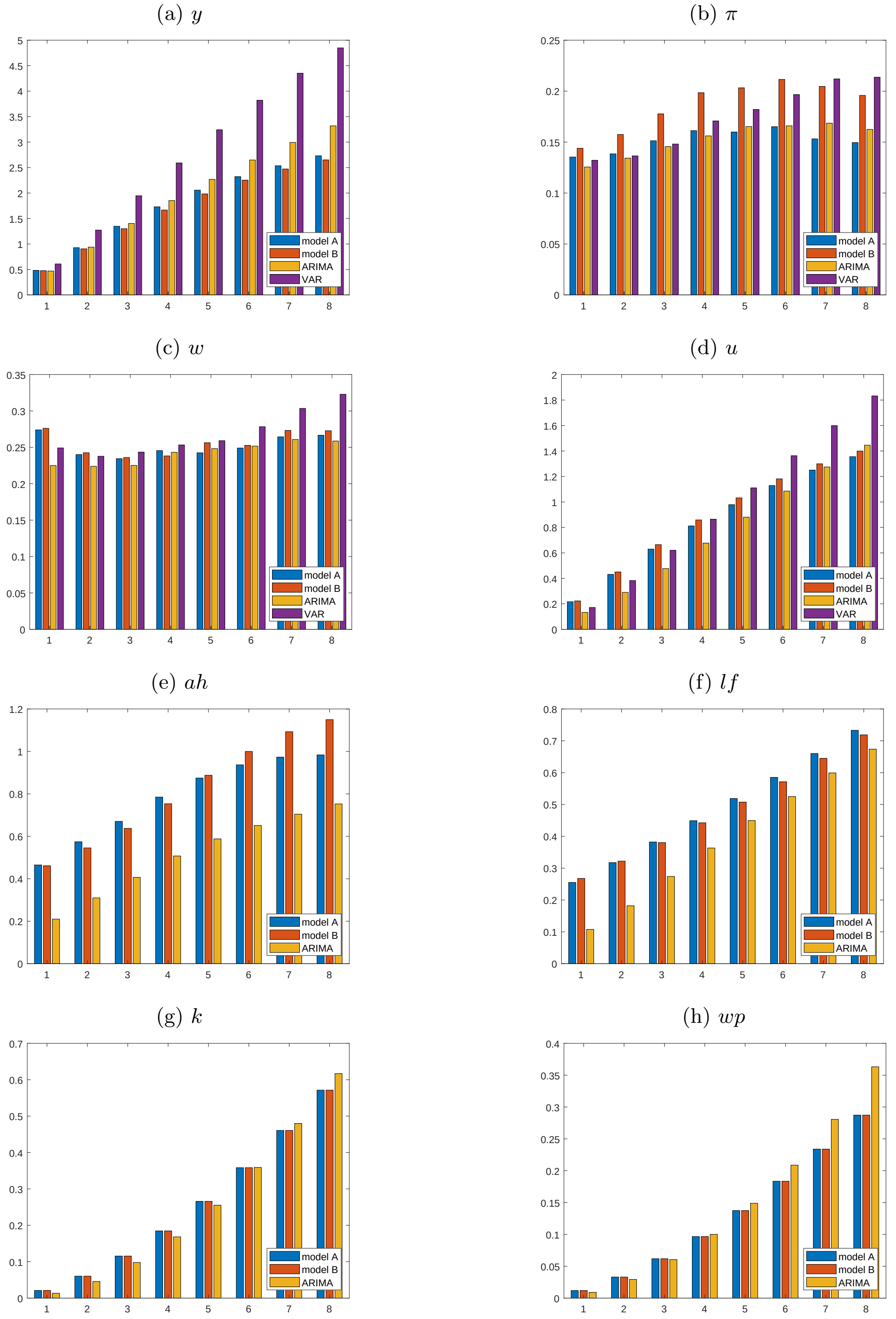
C Forecasting performance: detailed results

Figure C1: RMSFEs with fixed parameters



Note: The values on the horizontal axes refer to the length of the forecast horizons in quarters

Figure C2: RMSFEs with re-estimated parameters

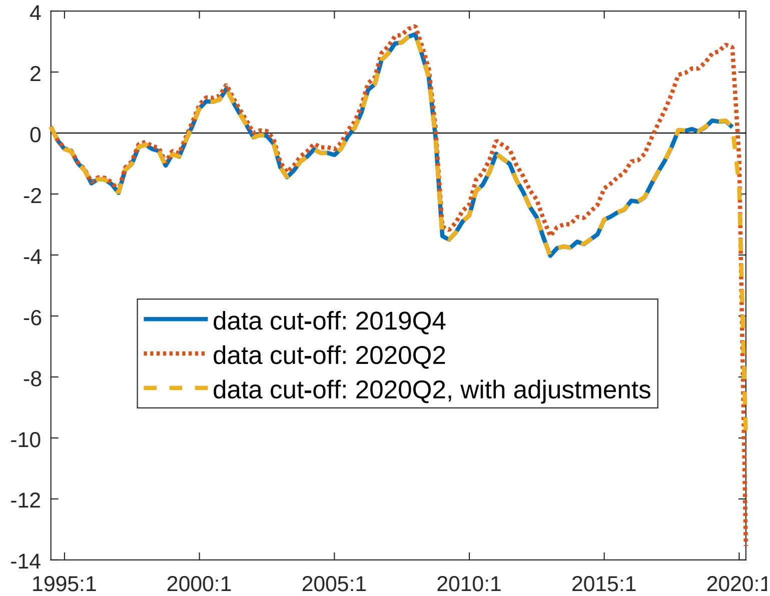


Note: The values on the horizontal axes refer to the length of the forecast horizons in quarters

D Output gap revisions and the Covid-19 shock

The Covid-19 pandemic and the related containment measures generated an unprecedentedly large shock to economic activity. Although the shock was fully exogenous - as opposed to the great financial crisis, when a buildup of financial imbalances preceded the recession - its sheer magnitude poses a problem for linear time-series models, such as the one introduced in this paper. From a trend-cycle decomposition perspective, the Covid-19 shock leads to an enormous prediction error in 2020 Q1 and Q2 which contaminates the the past state estimates if the Kalman smoother is applied. For example, by including the Covid-19 shock in the sample, but keeping the parameters of model A fixed at their levels estimated before the pandemic, the historical path of the output gap is revised up by 1 to 2 percentage points in each quarter over the past five years (see dotted red line vs solid blue line in Figure C1), generating a strongly positive output gap that diminishes the prediction error in the Covid-19 affected quarters. To avoid such an unwarranted revision to past states the adjustment similar to what was suggested by Lenza and Primiceri (2020) can be applied. In particular, it is possible to scale the covariance matrix of the state shocks - in an ad-hoc way - in the affected periods to accommodate the extraordinarily large shocks. Indeed, by applying a scale factor of 10 to the shock covariance matrix in the two Covid-19 affected quarters in the sample, the Kalman smoother does not revise the past states anymore (Figure C1).

Figure C1: Output gap revisions due to the Covid-19 shock (model A)



References

Adolfson, M., Laséen, S., Lindé, J., and Villani, M. (2007). Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-through. *Journal of International Economics*, 72(2):481–511.

Alichi, A., Avetisyan, H., Laxton, D., Mkhatrishvili, S., Nurbekyan, A., Torosyan, L., and Wang, H. (2019). Multivariate Filter Estimation of Potential Output for the United States: An Extension with Labor Market Hysteresis. IMF Working Papers 19/35, International Monetary Fund.

Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.

Andrle, M. and Plašil, M. (2018). Econometrics With System Priors. *Economics Letters*, 172(C):134–137.

Angeloni, I., Aucremanne, L., Ehrmann, M., Galí, J., Levin, A., and Smets, F. (2006). New evidence on inflation persistence and price stickiness in the euro area: Implications for macro modeling. *Journal of the European Economic Association*, 4(2/3):562–574.

Ball, L., Leigh, D., and Loungani, P. (2017). Okun's Law: Fit at 50? *Journal of Money, Credit and Banking*, 49(7):1413–1441.

Benes, J., Clinton, K., Garcia-Saltos, R., Johnson, M., Laxton, D., Manchev, P., and Matheson, T. (2010). Estimating Potential Output with a Multivariate Filter. IMF Working Papers 10/285, International Monetary Fund.

Benes, J., Johnston, M., and Plotnikov, S. (2020). IRIS toolbox release 20200613 (Macroeconomic modeling toolbox). Available at https://iris.igpmn.org/.

Blanchard, O., Cerutti, E., and Summers, L. H. (2015). Inflation and Activity: Two Explorations and Their Monetary Policy Implications. Working Paper Series WP15-19, Peterson Institute for International Economics.

Blanchard, O. and Summers, L. (1987). Hysteresis in Unemployment. *European Economic Review*, 31(1-2):288–295.

Bodnár, K. (2018). Labour Supply and Employment Growth. *Economic Bulletin Articles, European Central Bank*, 1.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

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

Camba-Mendez, G. and Rodriguez-Palenzuela, D. (2003). Assessment Criteria for Output Gap Estimates. *Economic Modelling*, 20(3):529–562.

Christiano, L., Trabandt, M., and Walentin, K. (2020). Involuntary Unemployment and the Business Cycle. *Review of Economic Dynamics*.

Christiano, L. J. and Fitzgerald, T. J. (2003). The Band Pass Filter. *International Economic Review*, 44(2):435–465.

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

Coibion, O., Gorodnichenko, Y., and Ulate, M. (2018). The Cyclical Sensitivity in Estimates of Potential Output. *Brookings Papers on Economic Activity*, 49(2 (Fall)):343–441.

Deoras, A. (2016). Fanchart - Visualize Percentiles of Time Series Data, MATLAB Central File Exchange.

Deroose, M., Rannenberg, A., and Wauters, J. (2019). Separating the Trend from the Cycle : The Debate on Euro Area Potential Output and Implications for Monetary Policy. *Economic Review*, pages 7–28.

Durbin, J. and Koopman, S. (2002). A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. *Biometrika*, 89(3):603–616.

Fagan, G., Henry, J., and Mestre, R. (2001). An Area-Wide Model (AWM) for the Euro Area. Working Paper Series 42, European Central Bank.

Fernald, J. G. and Wang, J. C. (2016). Why Has the Cyclicality of Productivity Changed? What Does It Mean? *Annual Review of Economics*, 8(1):465–496.

Geweke, J. (1999). Using simulation methods for bayesian econometric models: inference, development,and communication. *Econometric Reviews*, 18(1):1–73.

Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive Proposal Distribution for Random Walk Metropolis Algorithm. *Computational Statistics*, 14(3):375–395.

Harvey, A. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A. (2006). Forecasting with Unobserved Components Time Series Models. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1 of *Handbook of Economic Forecasting*, chapter 7, pages 327–412. Elsevier.

Harvey, A. and Trimbur, T. (2008). Trend Estimation and the Hodrick-Prescott Filter. *Journal of Japan Statistical Society*, pages 41–49.

Havik, K., Morrow, K. M., Orlandi, F., Planas, C., Raciborski, R., Roeger, W., Rossi, A., Thum, A. E., and Vandermeulen, V. (2014). The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps. European Economy - Economic Papers 535, DG ECFIN, European Commission.

Herbst, E. and Schorfheide, F. (2016). *Bayesian Estimation of DSGE Models*. Princeton University Press, 1st edition.

Hristov, A., Raciborski, R., and Vandermeulen, V. (2017). Assessment of the Plausibility of the Output Gap Estimates. Economic Brief 023, DG ECFIN, European Commission.

Jarociński, M. (2015). A Note on Implementing the Durbin and Koopman Simulation Smoother. *Computational Statistics & Data Analysis*, 91(C):1–3.

Jarociński, M. and Lenza, M. (2018). An Inflation-Predicting Measure of the Output Gap in the Euro Area. *Journal of Money, Credit and Banking*, 50(6):1189–1224.

Kaiser, R. and Maravall, A. (1999). Estimation of the Business Cycle: A Modified Hodrick-Prescott Filter. *Spanish Economic Review*, 1(2):175–206.

Koopman, S. J. and Harvey, A. (2003). Computing Bbservation Weights for Signal Extraction and Filtering. *Journal of Economic Dynamics and Control*, 27(7):1317–1333.

Lenza, M. and Primiceri, G. E. (2020). How to Estimate a VAR After March 2020. Working Paper Series 2461, European Central Bank.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21.

Nickel, C., Bobeica, E., Koester, G., Lis, E., and Porqueddu, M. (2019). Understanding Low Wage Growth in the Euro Area and European Countries. Occasional Paper Series 232, European Central Bank.

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

Pissarides, C. A. (1992). Loss of Skill During Unemployment and the Persistence of Employment Shocks. *The Quarterly Journal of Economics*, 107(4):1371–1391.

Primiceri, G. and Justiniano, A. (2009). Potential and Natural Output. 2009 Meeting Papers 25, Society for Economic Dynamics.

Proietti, T. and Musso, A. (2012). Growth Accounting for the Euro Area. *Empirical Economics*, 43(1):219–244.

Proietti, T., Musso, A., and Westermann, T. (2007). Estimating Potential Output and the Output Gap for the Euro Area: a Model-based Production Function Approach. *Empirical Economics*, 33(1).

Robert, C. P. and Casella, G. (2005). *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Stock, J. (1994). Unit Roots, Structural Breaks, and Trends. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, pages 2740–2843. Elsevier, Amsterdam.

Stock, J. H. and Watson, M. W. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, 39(s1):3–33.

Vetlov, I., Pisani, M., Hlédik, T., Jonsson, M., and Kucsera, H. (2011). Potential Output in DSGE Models. Working Paper Series 1351, European Central Bank.

Yellen, J. L. (2016). Macroeconomic Research After the Crisis. Remarks at "The Elusive 'Great' Recovery: Causes and Implications for Future Business Cycle Dynamics" 60th annual economic conference, Boston, Massachusetts.

Acknowledgements

I'm grateful to the members of a European System of Central Banks (ESCB) task force on potential output and an anonymous referee for their useful comments. The views expressed in this paper are those of the author and do not necessarily reflect those of the European Central Bank or the ESCB.

Máté Tóth
European Central Bank, Frankfurt am Main, Germany; email: mate.toth@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

This paper can be downloaded without charge from www.ecb.europa.eu, from the Social Science Research Network electronic library or from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website.

PDF ISBN 978-92-899-4523-3 ISSN 1725-2806 doi:10.2866/90703 QB-AR-21-014-EN-N