EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Marek Jarociński, Michele Lenza

An inflation-predicting measure of the output gap in the euro area

No 1966 / September 2016

Abstract

Using a small Bayesian dynamic factor model of the euro area we estimate the deviations of output from its trend that are consistent with the behavior of inflation. We label these deviations the output gap. In order to pin-down the features of the model, we evaluate the accuracy of real-time inflation forecasts from different model specifications. The version that forecasts inflation best implies that after the 2011 sovereign debt crisis the output gap in the euro area has been much larger than the official estimates. Versions featuring a secular-stagnation-like slowdown in trend growth, and hence a small output gap after 2011, do not adequately capture the inflation developments.

JEL Classification: C32, C53, E31, E32, E37

Keywords: Phillips curve, factor model, unobserved components, Bayesian estimation

Non-technical summary

The assessment of the output gap, i.e. the deviation of real economic activity from its potential, is an essential element for the implementation of economic policies aimed at stabilizing the economy. A large output gap usually calls for a demand stimulus, while slow trend growth is more conducive to supply-side policies. Many economists currently believe in a version of 'secular stagnation' hypothesis according to which developed economies, including the euro area, are facing a persistent slow-down of trend growth (see e.g. Gordon 2014). Moreover, some recent analyses carried out by international institutions suggest that the Great Recession and its aftermath resulted in a decline in potential output growth. Therefore, slack in the euro area economy is estimated to be smaller than what would follow from a scenario of stable potential growth. However, in practice, this assessment is far from straightforward, primarily because potential output is unobserved.

This paper presents an independent assessment of the euro area output gap, using a Bayesian dynamic factor model of core inflation (measured as the change in HICP excluding energy and food) and a set of real activity indicators. The model performs a trend/cycle decomposition of the real activity variables and core inflation. A single common factor captures common cyclical fluctuations and measures deviation of output from its trend, the output gap. Different modelling assumptions, such as different sets of real activity indicators and different models of trend components of the variables, lead to different estimates of the output gap. The model variants differ in three dimensions: the specification of the set of included real activity variables, the specification of the inflation trend (a simple univariate autoregressive process or proxied by Consensus long-term inflation forecasts), and, crucially, the form of the trends for the real activity variables (more or less adaptive to current economic conditions). The various combinations of the modelling assumptions just described leads to the estimation of seven different models. The resulting seven alternative estimates agree about the timing of peaks and troughs but they sometimes widely disagree about the level of the output gap. Such differences are economically relevant, for example, the output gap in 2014-2015 is, on average, about -2% according to some estimates, and -6% according to others. According to some models trend growth has not changed much over time and the recent sluggishness in real GDP in the euro area reflects a large output gap. This scenario calls for policies that stimulate aggregate demand in order to close the output gap. Other versions of the model, instead, imply that trend growth has strongly decelerated in the euro area and the output gap was closer to zero in 2014, a view associated with the secular stagnation hypothesis. This view suggests the need of supply side policies that would raise real GDP trend growth.

The paper conducts a real-time out-of-sample forecast evaluation in order to discriminate between the alternative measures of the output gap, at horizons ranging from 1 quarter to 1 year. In particular, the different variants of the dynamic factor model are ranked according to the precision of their out-of-sample, real-time forecasts of core inflation. The best model, according to this metric, provides accurate inflation forecasts over the last decade and implies that the output gap in the euro area has been about -6% in 2014 and 2015, on average. This estimate is considerably more negative than most official estimates, hovering between -2 and -3%, suggesting that the latter estimates may be subject to downward risks. Specifications of the model that estimate small output gaps and, consequently slow trend growth over recent years, forecast core inflation poorly over the evaluation sample and also recently. Hence, the reported results highlight that reconciling the above version of the secular stagnation hypothesis with the core inflation data remains a challenge.

1 Introduction

The assessment of how far an economy is from its potential is crucial for evaluating the inflationary pressures and determining the most appropriate economic policy mix. A negative output gap (the gap between the level of economic activity and its potential) usually calls for a demand stimulus, while slow trend growth requires supply-side policies. Our goal is to contribute to the assessment of economic conditions by developing a reliable measure of the euro area output gap that is relevant for inflation and can be used to interpret the economic developments in real time.

The approach we take in this paper is based on a dynamic factor model, estimated by means of Bayesian techniques, including a vector of euro area real activity indicators and core inflation. The long run behaviour of the variables is captured by variable-specific trends while their fluctuations at business cycle frequencies, around the trends, are captured by a common factor. This common factor is designed to coincide with the current deviation of real GDP from its trend and is, therefore, our measure of the output gap.

When setting up our model we face a number of modelling choices. Several observable variables can serve as imperfect measures of economic activity and inflation expectations. Which and how many observable variables should we include in our model? The properties of the trend processes crucially affect the decomposition into long run trend and business cycle. How to specify the trend processes? We find that different reasonable combinations of these model features lead to euro area output gap estimates that broadly agree about the timing of peaks and troughs, but widely disagree about the level of the output gap. This raises the issue of which of these estimates is the most accurate. However, because it is an unobservable variable, no empirical validation can be directly conducted on the output gap itself.

In order to discriminate among different measures of the output gap we rely on a Phillips curve-type relationship linking the output gap with inflation. This relationship has been the cornerstone of most empirical work on inflation forecasting (see Stock and Watson, 1999 for a prominent example, and Faust and Wright, 2013 for a recent survey). Our model validation strategy relies on the idea that a policy-relevant measure of the output gap should send an accurate signal about future inflation in real-time. Therefore, we rank the different variants of our dynamic factor model according to the precision of their out-of-sample, real-time forecasts of core inflation. Our validation exercise exploits a database for the euro area that tracks the real-time information set of the European Central Bank (see Giannone et al., 2012).

We find that the best output gap estimates are extracted from a relatively large set of observable variables, with relatively inflexible trend processes, and that it is useful to link trend inflation to long term inflation expectations. The resulting forecasts track actual inflation quite well, both before the 2008 crisis, and since its onset, when they correctly capture the fall in inflation. The output gap in the crisis is large and over the period 2014-2015 it reaches -6% of euro area GDP, a value twice as negative as the publicly available estimates of international institutions. Moreover, we find that specifications of our model that produce slow trend growth and, consequently, small output gaps, forecast core inflation poorly. This finding is particularly interesting in light of the recent debate about the 'secular stagnation' hypothesis, according to which the weak growth in the advanced economies after the crisis reflects the slowdown of the trend growth (see e.g. Summers, 2013; Gordon, 2014). This hypothesis is widely debated and many other studies conclude instead that this recent weak growth reflects cyclical, although persistent, sources of fluctuations (see Blanchard, 2015 for a recent survey of the arguments of the two sides). Our results highlight that reconciling a slowdown in trend growth with the data on euro area core inflation remains a challenge.

The real-time perspective we take in our paper also allows us to meaningfully study the reliability of the end-of-sample output gap estimates, which is crucial if they are to be of use for policy. In their influential paper, Orphanides and van Norden (2002) show that ex post revisions of the real-time end-of-sample output gap estimates are of the same order of magnitude as the output gap itself, rendering it virtually useless in practice. We find that small models, similar to the ones studied by Orphanides and van Norden (2002), are indeed as unreliable as they report. However, we find that the output gap estimates from our best model, which is much larger, are revised much less as data accumulate, so that they turn out to be reasonably reliable in real time.

This paper is related to a large literature on output gap and Phillips curve estimation with unobserved components models. The small-scale Phillips curve model of Kuttner (1994) initiated this literature. Planas et al. (2008) estimate a Bayesian version of Kuttner's model and we build on their priors. Similarly as Baştürk et al. (2014) we use non-filtered data and pay much attention to modelling their low frequency behavior. We confirm the finding of Valle e Azevedo et al. (2006) and Basistha and Startz (2008) that using multiple real activity indicators increases the reliability of output gap estimates. Following Valle e Azevedo et al. (2006) and D'Agostino et al. (2015) our model accounts for the presence of not only coincident, but also leading and lagging indicators, although we use a different parameterization. Finally, Faust and Wright (2013) and Clark and Doh (2014) document the advantages of relating trend inflation to data on long-term inflation expectations. Indeed, we find that relating trend inflation to long-term inflation expectations is a crucial ingredient of a successful output gap model in our application. We build on all this literature, and the distinguishing feature of our paper is that we use the Phillips curve-type relation in the real-time out-of-sample context to select the model and thus pin down the estimate of the unobserved output gap.

The rest of the paper is organized as follows. Section 2 briefly describes the real-time database. Section 3 describes the model and its estimation. Section 4 reports the empirical results. Section 5 concludes.

2 Data

In this paper, we adopt a fully real-time data perspective. The macro-econometric literature has emphasized the relevance of the real-time data uncertainty about the output gap (Orphanides and van Norden, 2002). Our data source is the euro area real-time database described in Giannone et al. (2012). The frequency of our dataset is quarterly. For the variables that are reported at the monthly frequency, we take quarterly averages. All variables are seasonally adjusted in real-time.

The first block of our dataset consists of seven indicators of real economic activity: real GDP (y_t^1), real private investment (y_t^2), real imports (y_t^3), real export (y_t^4), unemployment rate (y_t^5), consumer confidence (y_t^6) and capacity utilization (y_t^7). The first four variables are in log levels, the remaining three in levels.

Our measure of prices is the Harmonized Index of Consumer Prices (HICP) excluding energy and unprocessed food prices. The log of this index is denoted p_t and the inflation variable that enters the econometric model is $\pi_t = 400(p_t - p_{t-1})$.

We also use the 6-to-10-year ahead inflation expectations (π_t^e) for the euro area from Consensus Economics. Since 1989, Consensus Economics collects and publicly releases, every April and October, 6-to-10-year inflation forecasts of G-7 countries. Starting in 2003 Consensus Economics reports the forecasts for the euro area as a whole. Before 2003, we compute the 6-to-10-year inflation expectations for the euro area by weighing the forecasts for Germany, France and Italy according to their GDP levels. We assign the April release to the second quarter and the October release to the fourth quarter of the respective year. Pre-1989 inflation expectations and those of the first and third quarter of each year are treated as missing data.

For each variable, we collect the 54 real-time data vintages released in the beginning of the third month of each quarter from 2002Q3 to the 2015Q4.[1] Consequently, for the last quarter of each real-time sample we only observe capacity utilization (which comes from a survey) and inflation expectations (in the second and fourth quarter of the year, otherwise the last quarter is also missing), while for the other indicators the last available release refers to the previous quarter (GDP, inflation, unemployment, consumer confidence) or to two quarters earlier (investment, exports and imports). Hence, our real-time database is characterized by a “ragged edge”, i.e. it has missing values at the end of the sample, in addition to the missing values of inflation expectations in half of the quarters.

The sample starts in 1985Q1 in each vintage. The 30 observations from 1985Q1 to 1992Q2 are used as a training sample, to inform our prior. Observations starting from 1992Q3 are used for the estimation.[2]

[1]The database of Giannone et al. (2012) collects the data vintages reported each month in the ECB Economic Bulletin (formerly Monthly Bulletin), it is regularly updated and publicly available in the ECB Statistical Data Warehouse.

[2]For more information on the database, see data appendix at the end of the paper.

3 Econometric model

We use an unobserved components model (a dynamic factor model) to estimate the output gap and to forecast inflation. The model relates the observed variables – real activity indicators $y_t^1, ..., y_t^N$, inflation π_t, and inflation expectations π_t^e – to unobservable state variables. The set of unobservable state variables includes the common stationary component g_t (which, as we explain below, is our measure of the output gap), trends of real activity indicators, $w_t^1, ..., w_t^N$, and trend inflation z_t.

The observation equations of the model are

$$y_t^n = b^n(L)g_t + w_t^n + \varepsilon_t^n, \qquad \text{for } n = 1, \ldots, N, \tag{1a}$$

$$\pi_t = a(L)g_t + z_t + \varepsilon_t^\pi, \tag{1b}$$

$$\pi_t^e = c + z_t + \varepsilon_t^e, \tag{1c}$$

where $\varepsilon_t^n, \varepsilon_t^\pi, \varepsilon_t^e$ are independent Gaussian errors, $b^n(L)$ and $a(L)$ are polynomials in the lag operator L and c is a constant term.

The state equations in the baseline version of the model are

$$g_t = \phi_1 \, g_{t-1} + \phi_2 \, g_{t-2} + \eta_t^g, \tag{2a}$$

$$w_t^n = d^n + w_{t-1}^n + \eta_t^n \qquad \text{for } n = 1, \ldots, N, \tag{2b}$$

$$z_t = d^z + f z_{t-1} + \eta_t^z, \tag{2c}$$

where $\eta_t^g, \eta_t^n, \eta_t^z$ are independent Gaussian errors, and ϕ_1, ϕ_2, d^n, d^z and f are coefficients.

The first observation equation (1a) relates the n-th real activity variable y_t^n to a variable specific trend w_t^n and to g_t, a common factor. The latter may enter the equations both with a lead and with a lag, as $b^n(L)$ are polynomials with both negative and positive powers of L.[3] In so doing, we accommodate for the presence of both contemporaneous, lagging and leading indicators in the vector of real activity variables. The first variable, y_t^1, is the log of the real GDP and for this variable we restrict the coefficient of g_t to be 1, the coefficients

[3]See Valle e Azevedo et al. (2006) and D'Agostino et al. (2015) on the importance of capturing dynamic heterogeneity across variables to appropriately extract the common cyclical features in the variables.

of lagged and future g to be zero, and the shock variance to zero, so this equation reads $y_t^1 = g_t + w_t^1$. This restriction identifies g_t as the current output gap (deviation of real GDP, y_t^1 from its trend, w_t^1) and ensures that it is expressed in percent of real GDP.[4]

The second observation equation (1b), the Phillips Curve, relates inflation π_t to the current and lagged output gap g_t and to trend inflation z_t. Similar equations estimated in the literature often include two additional features, which we consciously omit here: variables reflecting external cost-push shocks, and stochastic volatility (time-varying variance of the shocks). None of these features seems to matter for our measure of inflation, which is based on HICP excluding energy and unprocessed food in the sample 1992-2015. First, we have tried including the commodity price index, the oil price and the exchange rate, and we found that these typical cost-push variables do not contribute much to modelling our core inflation measure. Appendix D reports the details. Second, we have estimated the Stock and Watson (2007) unobserved components/stochastic volatility (UCSV) model of inflation and tested for the presence of stochastic volatility following Chan (2016). While the evidence for stochastic volatility is sometimes there for much longer samples and for headline measures of inflation, in our sample stochastic volatility is not needed. Appendix E reports the details. This evidence is based on Stock and Watson's small model, but we conjecture that the lessons extend to a richer model like our baseline model.

The third observation equation (1c) relates trend inflation z_t to long term inflation expectations π_t^e, as advocated by e.g. Faust and Wright (2013) and Clark and Doh (2014). Cogley et al. (2010) provide an economic rationale for trend inflation. The constant term c introduces a wedge between trend inflation and inflation expectations. This wedge accounts for the fact that inflation expectations π_t^e and inflation π_t refer to different concepts of inflation (headline HICP and HICP excluding energy and unprocessed food, which tends to be lower), as well as for any systematic biases in inflation expectations such as those detected by Chan et al. (2015).

The state equation (2a) specifies an AR(2) process for g, while the state equation (2b) specifies the trend of each real activity variable to be a random walk with drift (RW). In

[4] In addition to facilitating the economic interpretation of g, this restriction is crucial to avoid the usual scale indeterminacy of factor models as well as to fix the reference phase.

alternative versions of the model we replace the random walk trends by

$$\text{i.i.d. trend,} \qquad w_t^n = d^n + \eta_t^n, \tag{3}$$

$$\text{Integrated Random Walk (IRW)} \qquad w_t^n = \delta_t^n + w_{t-1}^n, \tag{4a}$$

$$\delta_t^n = \delta_{t-1}^n + \zeta_t^n, \tag{4b}$$

$$\text{and Local Linear Trend (LLT)} \qquad w_t^n = \delta_t^n + w_{t-1}^n + \eta_t^n, \tag{5a}$$

$$\delta_t^n = \delta_{t-1}^n + \zeta_t^n. \tag{5b}$$

It is worth noting that the IRW and LLT processes are both more flexible than the random walk with drift, as they allow the drift to change over time.[5]

Each real activity variable has its own trend. An alternative would be to specify common trends for some of the related real activity variables, but we do not pursue this possibility here.

Finally, the state equation (2c) specifies a stationary AR(1) process for the trend inflation.

3.1 Priors

Our prior choice is based on the use of a training sample and an adaptation of the approaches popularized by the literature on time-invariant and time-varying parameter Bayesian vector autoregressions. Here we only sketch the most relevant aspects, while the full details are reported in Appendix B.

In practice, we center our priors around the simple model in which each observable variable is a sum of a random walk trend and an i.i.d. noise. Based on the training sample, we calibrate the prior variances of the shocks to trend and noise so that they each explain one-half of the variance of the first difference of the variable. The loadings of each variable on the output gap are centered at zero, with variances scaled as in the Minnesota prior for

[5] We experiment with the flexibility of the trend process, while maintaining the assumption that the trend and cycle errors are uncorrelated. Morley et al. (2003) show that when these errors are modeled as correlated, the random walk trend explains more and cycle less of the variation in output, hence this has a qualitatively similar effect to specifying a more flexible trend process.

vector autoregressions. We introduce a subjective prior about the properties of the output gap process, which captures the stylized facts on the periodicity and persistence of the business cycles.

The functional form of the priors about shock variances is inverted gamma and all the remaining priors are Gaussian. These functional forms ensure a convenient and fast computation of the posterior with the Gibbs sampler.

3.2 Estimation

In the Gibbs sampler we draw the parameters ($b^n(L)$ and d^n for $n = 1, ..., N$, $a(L)$, c, ϕ_1, ϕ_2, d^z, f, and all the shock variances) conditionally on the unobserved states (g_t, w_t^n for $n = 1, ..., N$, z_t, all for $t = 1, ..., T$), and then draw the states conditionally on the parameters. The conditional posteriors of the parameters are Gaussian and inverted gamma. The conditional posterior of the states is Gaussian, and we draw from it using the simulation smoother of Durbin and Koopman (2002), implemented as explained in Jarociński (2015). To compute each posterior we generate 250,000 draws with this Gibbs sampler, out of which we discard the first 50,000. We assess the convergence of the Gibbs sampler using the Geweke (1992) diagnostics, see Appendix C.

4 Empirical results

4.1 Model specifications

We estimate seven variants of the model. Comparing these variants helps us to understand the role of various features of the model. In particular, the models differ in three dimensions: the real activity variables included in the model, the inclusion of long term inflation expectations, and the functional form of the trends of real activity variables. Table 1 provides an overview.

Model 1 includes only inflation and real GDP, the minimal set of variables to extract the output gap and forecast inflation. Models 2 extends Model 1 by including long term inflation expectations that pin down trend inflation. Model 3 extends Model 1 by including all the seven indicators of economic activity. Models 4 to 7 feature both long term inflation

Table 1 – Model specifications.

	trend y	trend π related to	Variables π_t	π_t^e	y_t^1	y_t^2	y_t^3	y_t^4	y_t^5	y_t^6	y_t^7
Model 1	RW	-	x		x						
Model 2	RW	π_t^e	x	x	x						
Model 3	RW	-	x		x	x	x	x	x	x	x
Model 4	RW	π_t^e	x	x	x	x	x	x	x	x	x
Model 5	RW or i.i.d.	π_t^e	x	x	x	x	x	x	x	x	x
Model 6	IRW	π_t^e	x	x	x	x	x	x	x	x	x
Model 7	LLT	π_t^e	x	x	x	x	x	x	x	x	x

Note: The variables used to estimate each model are indicated with an x in the columns 4 to 12. π_t is the quarterly percentage change in HICP excluding energy and unprocessed food; π_t^e is the long-term inflation expectation from Consensus Economics; y_t^1 is real GDP; y_t^2: real private investment; y_t^3: real imports; y_t^4: real exports; y_t^5: unemployment rate; y_t^6: consumer confidence; y_t^7: capacity utilization.

expectations and all the seven indicators of real activity.

In Models 1 to 4 the trends of the real activity variables are modeled as random walks with drift. By contrast, trends in Model 5 are more rigid, and in Models 6 and 7 they are more flexible. In particular, in Model 5 the trends of the three a priori stationary real activity variables (unemployment rate, consumer confidence and capacity utilization) are modeled as i.i.d. processes with a constant mean. In Models 6 and 7, the trends of all real activity variables are modeled as integrated random walks and as local linear trends, respectively.

4.2 Output gap estimates on the last vintage of the data

We start by estimating each of the seven models on the longest sample, 1992Q3 to 2015Q4. Figure 1 plots the point estimates (posterior medians) of the output gap over time. This figure shows that the peaks and troughs of the output gap estimates typically coincide across models. However, the results also highlight that it is important to discriminate among the model features we have discussed above, because different combinations of trend specifications and observables lead to substantial disagreement about the size of the output gap. For example, in the years 2014-2015 the estimates of the output gap range from less than -2 to -6 percent of GDP.

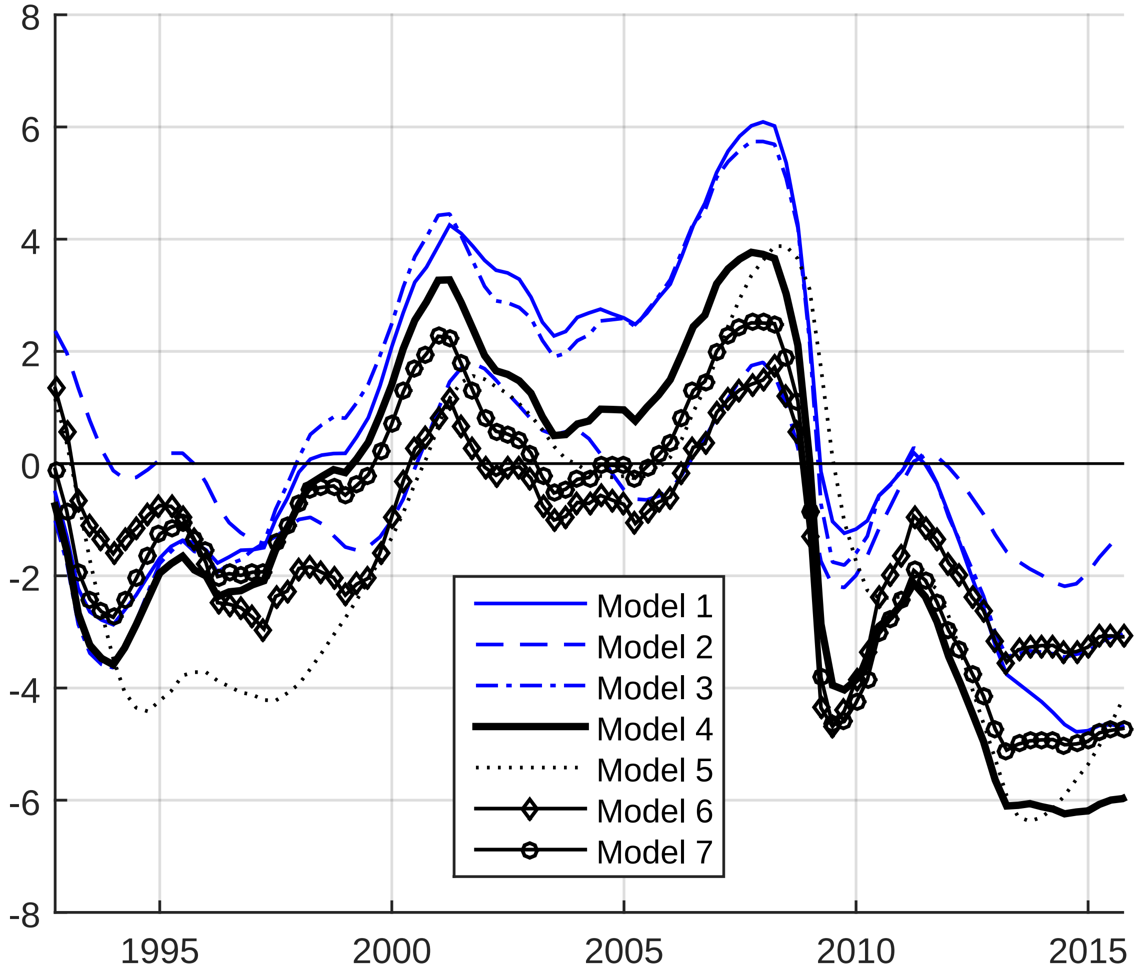


Figure 1 – Point estimates (posterior medians) of the output gap from Models 1 to 7

4.3 Forecasting results

In this subsection, we discriminate among the different measures of output gap by studying the associated real-time out-of-sample forecasts of inflation. The design of this empirical validation exercise is as follows. We re-estimate each of the seven models over 54 expanding samples of our real-time data and, for each sample, we forecast inflation up to one year ahead. Our target measure of inflation for horizon h is the annualized rate of change in consumer prices, $400/h\,(p_{t+h} - p_t)$, where p_t is the log-level of the price index. We compute the target inflation rate using the latest available data, i.e. those available as of 2015Q4.

The first estimation sample (data available on 2002Q3) spans the period 1992Q3 - 2002Q3 and the last (data available on 2015Q4) 1992Q3 - 2015Q4. As explained in the data section, we have a 'ragged edge' due to the different timeliness of data releases. In particular, our information set in quarter t, denoted $\mathcal{I}(t)$, contains data on capacity utilization up to t, but e.g. prices only up to $t-1$. Therefore, the quantity we forecast is $400/h\,(p_{t-1+h} - p_{t-1})|\mathcal{I}(t)$. We evaluate both the point and the density forecasts.

We start with the evaluation of point forecasts. Panel A of Table 2 reports the mean squared error (MSE) of the nowcast of inflation $400\,(p_t - p_{t-1})|\mathcal{I}(t)$, one quarter ahead forecast $200\,(p_{t+1} - p_{t-1})|\mathcal{I}(t)$ and, one year ahead forecast, $100\,(p_{t+3} - p_{t-1})|\mathcal{I}(t)$. Our point forecasts are the medians of the posterior predictive densities. For comparison, we report also the MSE of a simple benchmark model, the unobserved components - stochastic volatility (UCSV) model of Stock and Watson (2007).[6] It would be interesting per se to study the performance of other simple benchmarks too, but here we are mainly interested in ranking our output gap models and the choice of the benchmark does not matter for this ranking.[7]

Panel A of Table 2 shows that the Phillips curve inflation forecasts (inflation forecasts using information on real activity) can outperform our simple benchmark in this sample, but the specification of the model matters crucially for the forecasting performance. First, the models including the full set of real economy variables (Models 3 to 7) are generally performing better at longer horizons than those with real GDP only. This suggests that the larger information set allows the extraction of a more timely and precise measure of the latent output gap. The second important result relates to the role of long term inflation expectations for the estimation of trend inflation. Generally, the models including the measure of long term inflation expectations provide better forecasts of inflation (particularly at the one year horizon) than those with a comparable set of real activity variables and excluding inflation expectations. In particular, Model 3, which does not include long term inflation expectations, is dominated by Models 4 to 7, which do include the expectations.

[6] We have estimated the USCV model using the codes accompanying Chan (2016). The UCSV model does not forecast so well in our sample because our measure of inflation is much less volatile than the inflation measures for which the UCSV model typically works well. In particular, our measure of inflation is the quarterly, seasonally adjusted HICP excluding prices and unprocessed food. Since our data is so smooth, the estimated trend of the UCSV model is close to the actual inflation. As a result, the forecasts of inflation are close to the last period value of inflation, and this is not a good forecast in our sample.

[7] We have also tried the random walk with drift for the price level p_t, another standard simple benchmark that is difficult to beat in the euro area *prior to the crisis*, see e.g. Diron and Mojon (2008), Fischer et al. (2009), Giannone et al. (2014), but it performed worse than the UCSV model.

Table 2 – Inflation forecasts: real time performance.

	A. Mean squared error of point forecast			B. Mean log predictive density score		
	$h=1$	$h=2$	$h=4$	$h=1$	$h=2$	$h=4$
Model 1	0.17	0.17	0.22	**-0.55**	-0.50	-0.58
Model 2	0.18	0.18	0.24	**-0.54**	-0.56	-0.72
Model 3	0.18	0.17	0.22	**-0.57**	-0.53	-0.63
Model 4	**0.16**	**0.13**	**0.13**	-0.58	**-0.45**	**-0.50**
Model 5	0.23	0.19	0.17	-0.69	-0.61	-0.55
Model 6	0.18	0.17	0.21	-0.63	-0.52	-0.61
Model 7	0.18	0.15	0.14	-0.64	-0.52	-0.64
Simple benchmark	0.22	0.21	0.27	-0.77	-0.77	-0.89

Note: Point forecast is the median of the predictive density. Evaluation sample: 2002Q3-2015Q4. The forecast for $h = 1$ is a nowcast.

Finally, allowing for the more flexible trend representations embedded in Models 6 and 7 is not particularly helpful in terms of forecasting accuracy, as these models are less accurate than Model 4. Summing up, the model delivering the best forecasting performance is our Model 4, which includes the whole set of real economy variables, the measure of inflation expectations to inform the trend inflation and a parsimonious random walk representation for the trends in the real economy variables. For this reason, we take Model 4 as our baseline model.[8] Panel B of Table 2 corroborates these findings and complements the picture by reporting the log scores of predictive densities. Model 4 outperforms other specifications also when we evaluate the accuracy of the full predictive distribution. The exception is the nowcast, where Models 1 to 3 outperform the larger models. This is because the variances of the forecasts from Models 1, 2 and 3 tend to be larger. Larger variances help to mitigate the effects of forecast errors on the log score at short horizons, where errors tend to be small. Instead, Models 4, 5, 6 and 7 have generally tighter predictive distributions, which are strongly penalized in terms of log scores when the median of the distribution is ‘far’ from observed inflation. Overall, Model 4 provides the best compromise, as it performs well

[8]Notice that Model 7, in terms of point forecasts, is closer to Model 4 than the other models. However, it turns out that the result is due to the fact that the estimate of the trend economic activity in Model 7 are quite similar to those from Model 4 and, hence, the result further confirms the view that a relatively parsimonious specification for the trend of economic activity is preferred by the data.

both at the short and at the long horizons and both in terms of point and density forecasts.

Figure 2 presents one year ahead predictive densities of inflation along with the actual inflation. The solid line is the observed annual inflation, final vintage, i.e. at time t it represents $100(p_t - p_{t-4})|\mathcal{I}(2015Q4)$. The blue lines and shaded regions indicate the 50th, 16th and 84th quantiles of the real time predictive density of inflation one year ahead, i.e. the density of $100(p_t - p_{t-4})|\mathcal{I}(t-3)$. The figure clearly shows that that the one-year-ahead forecasts of inflation become much less volatile when (i) we use multiple indicators of real activity and (ii) we relate trend inflation to long term inflation expectations. In fact, the inflation forecasts are much more volatile in Models 1 to 3 than in Models 4 to 7.

4.4 How robust are the estimates of the output gap in real-time?

The issue of robustness of the end-of-sample estimates of the output gap has attracted much attention since Orphanides and van Norden (2002), who argue that revisions to real-time end-of-sample output gaps are of the same order of magnitude as the output gaps themselves, rendering the output gaps virtually useless for a policy maker. Our framework allows us to study the robustness of our estimates of output gap in real time.

To summarize the real time revisions of the output gap we compute the envelope of the 16th and 84th percentiles related to our 54 real time posteriors. More in details, at each date we have up to 54 sets of posterior quantiles of the output gap, obtained with our 54 real time samples. The envelope percentiles are computed, at each date, as the lowest of the available 16th percentiles, and the highest of the available 84th percentiles. Figure 3 plots these envelope percentiles over time, along with the percentiles obtained in the last sample, which spans the period 1992Q3-2015Q4.

This figure confirms the validity of Orphanides and van Norden (2002) concerns. In Model 1, which is similar to the models they study, output gap revisions are indeed of the similar order of magnitude as the output gap itself, and hence the envelope includes zero in almost all the periods. However, in larger models the envelopes are narrower. In particular, the lessons about the output gap coming from Model 4 are reasonably robust in real time.[9] Hence, in some models robustness is indeed a serious concern. However, output

[9]Mertens (2014) shows that adding stochastic volatility to a small model like Model 1 improves the real-

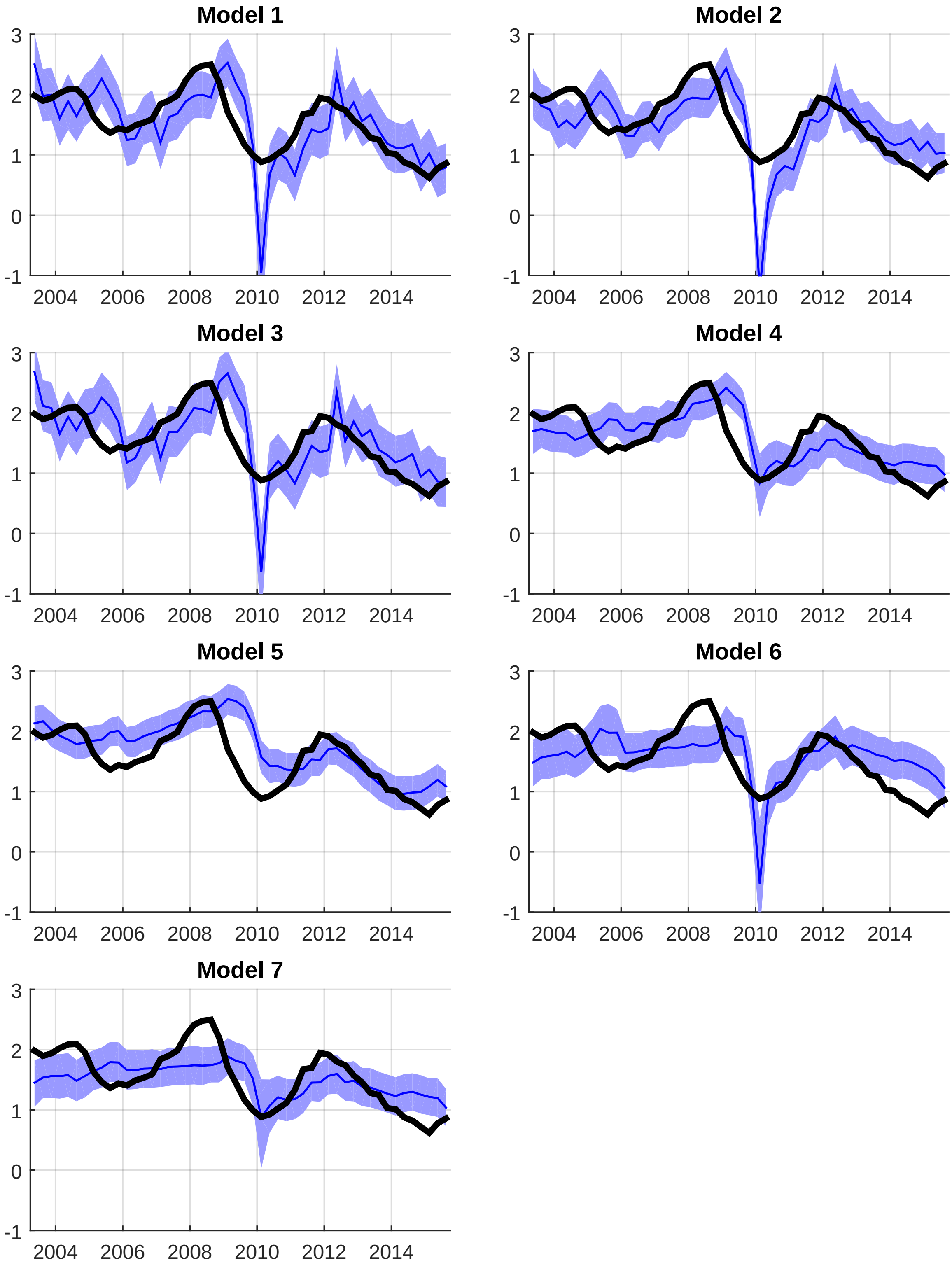


Figure 2 – One year ahead forecasts of inflation (blue line: median, blue shaded area: percentile 16 to 84) and actual inflation (black line)

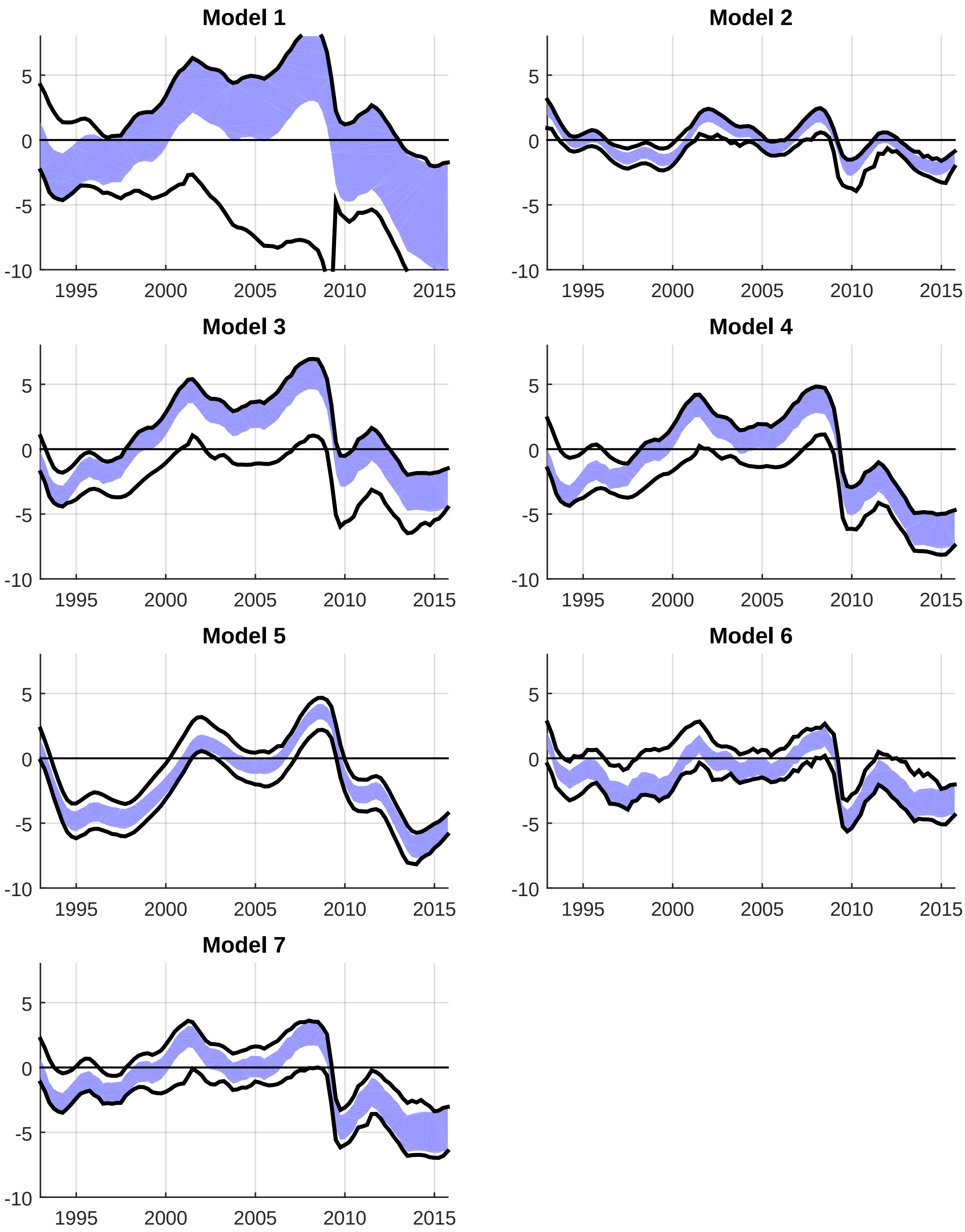


Figure 3 – 16th and 84th percentiles of the output gap: envelope of all the real time samples (black line) and the last sample (blue shaded area)

gap estimates from our best performing model, Model 4, turn out to be quite robust in real time.

4.5 The output gap after the double dip recession

According to our best performing models, in the years 2014-15 we observe the largest (negative) output gap in the history of the euro area. An important policy implication of this fact is that, currently, a demand stimulus that would close this output gap is more urgent than structural reforms. This view is, however, not consensual, since many believe that a crucial problem facing the euro area is that trend output growth has stalled. The latter view (which could be seen as a more 'conjunctural' variant of the 'secular stagnation' hypothesis) would imply also that, currently, there is not much of a gap between trend and actual output, in the euro area, and that structural reforms rather than demand stimulus are needed to revive output.

Figure 4 presents a comparison of the outcomes from Model 4 and Model 6. These two models represent, among our models with the full set of observables,[10] the two polar views described above. In particular, Model 4 is consistent with the view that trend has not changed much and the low observed output is a result of a large output gap while Model 6 is consistent with the view that trend output growth has markedly decreased since the beginning of the financial crisis. The official output gap estimates by the IMF and the European Commission, also shown in this figure, are close to those obtained by Model 6. In the first recession, in 2009, Model 4 produces a similar assessment of the output gap as the IMF and the European Commission. However, in the second recession these assessments diverge.

Figure 4 makes it clear that the properties of the trend output are crucial. In Model 4, the growth rate of trend output hardly changes. Then, when economic growth stalls, as it has since the beginning of the Great Recession in the euro area, the output gap opens ever wider. By contrast, in Model 6 the growth rate of trend output falls sharply, and trend

time reliability of output gap estimates. We leave it for future research to study whether stochastic volatility could make a difference in our context.

[10] Here, we restrict ourselves to the set of large models because we are commenting on relatively recent developments and, in the previous sub-section, we showed that large models provide more reliable assessments of output gap over time.

output tracks the developments in actual output more closely. As a consequence, according to Model 6 and to the official estimates the output gap in the second dip has been smaller in absolute value than it was in the first dip.

This paper shows that the view that trend output growth has strongly slowed down in the euro area produces worse inflation forecasts than the competing view embodied in Model 4, because it is hard to reconcile the persistently low inflation with small output gaps. We find that Model 4 (as well as Models 5 and 7, which also estimate a large output gap and a relatively unchanged potential output growth) clearly dominates Model 6 which, among other things, persistently overpredicts inflation after 2012. Incidentally, also the official inflation forecasts of the IMF and the European Commission, whose output gap estimates resemble a lot those of Model 6, were excessively high after 2012.[11]

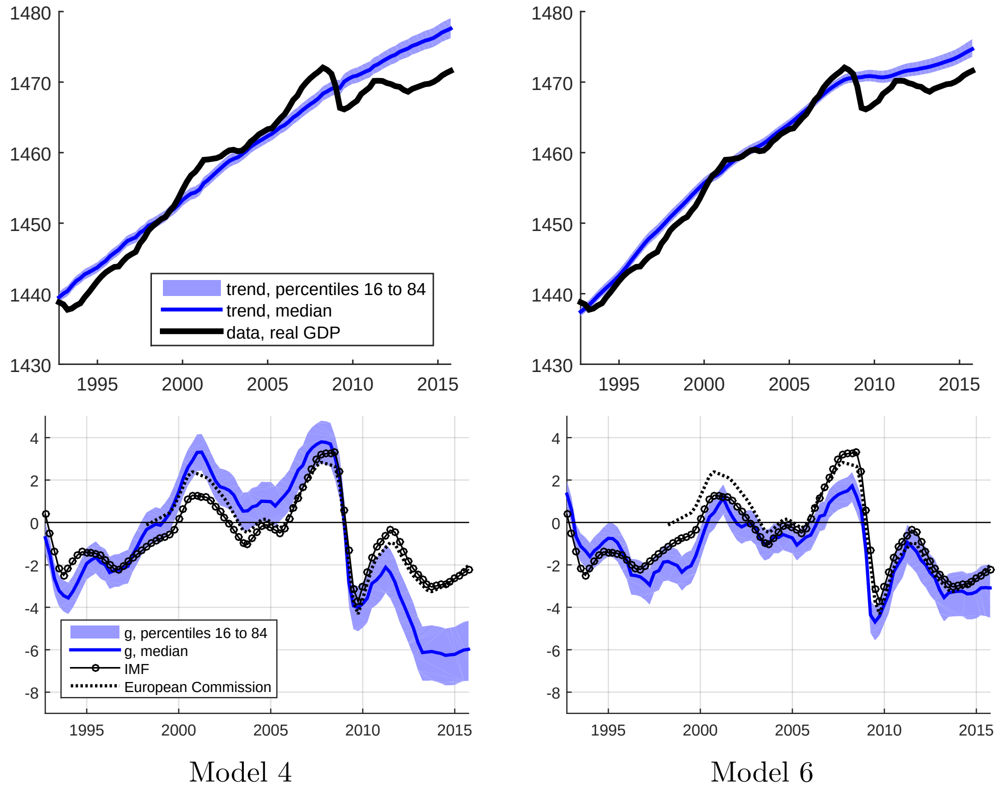


Figure 4 – Trend output and output gap according to Models 4, 6, IMF and European Commission

[11]Our framework is specifically geared to forecast inflation. However, the model can also produce forecasts of the real activity variables. See the Appendix for an evaluation of the marginal likelihoods.

5 Conclusions

We estimate the output gap in the euro area with several specifications of a Bayesian dynamic factor model. We find that while alternative specifications agree about the timing of peaks and troughs, they disagree about the size of the output gap. We find that the real-time inflation forecasts generated by these models improve when we include multiple real activity indicators, when we relate trend inflation to long term inflation expectations, and when we model real activity trend components as random walks, instead of either more or less flexible processes.

Our estimate of the output gap has three appealing features from the point of view of policy makers: it is a measure of the slack of the economy, it helps to forecast inflation, and it is quite reliable in real time. Our estimate suggests that after the second dip of the recent recession the output gap is even larger than it was in the first dip, and allows us to correctly predict falling inflation since 2012.

Appendix

Appendix A Data appendix

Table A.1 reports for each variable the definition (column 1), mnemonic (column 2), transformation (column 3), the latest period of availability in the data vintage dated t (column 4), the data source (column 5) and the part of the training sample for which we back-dated the series using the Area Wide Model (AWM) database (Fagan et al., 2001) (column 6).

Table A.1 – The description of the variables

Variable name	Symbol	Transf.	Availability in vintage t	Source	Backdating from AWM
HICP excl. energy and unprocessed food	p	log-diff (π)	$t-1$	Euro area RTD	85Q1-89Q4
Real GDP	y^1	log	$t-1$	Euro area RTD	85Q1-90Q4
Real private investment	y^2	log	$t-1$	Euro area RTD	85Q1-90Q4
Real imports	y^3	log	$t-2$	Euro area RTD	85Q1-90Q4
Real exports	y^4	log	$t-2$	Euro area RTD	85Q1-90Q4
Unemployment rate	y^5	none	$t-1$	Euro area RTD	85Q1-90Q2
Consumer confidence	y^6	none	$t-1$	Euro area RTD	None
Capacity utilization	y^7	none	t	Euro area RTD	None
Inflation expectations	π^e	none	t or $t-1$	Consensus Economics	None

Our training sample goes back to 1985Q1, but for some variables the euro area RTD data start only in 1990. In those cases, we extend the series from the first vintage back in time using the growth rates of the respective series from the Area Wide Model (AWM) database. Note that this back-dating affects only the training sample, 1985Q1-1992Q2. The post-1992Q2 samples used for the main analysis come exclusively from the real-time database.

Appendix B The priors

We first describe the priors in the baseline model and then explain how the priors differ in the alternative versions of the model.

The first step in our strategy for prior selection is to compute the mean and variance of the first difference of each observable variable in the training sample. Let T^{tr} denote the size of the training sample. For each variable $v \in \{y_t^1, ..., y_t^N, \pi_t, \pi_t^e\}$ we compute the mean $\tilde{\delta}_v = \frac{1}{T^{tr}-1}\sum_{t=2}^{T^{tr}} \Delta v_t$ and variance $\tilde{\sigma}_v^2 = \frac{1}{T^{tr}-1}\sum_{t=2}^{T^{tr}} (\Delta v_t - \delta_v)^2$.

Coefficients of the observation equations. The coefficients $b^n(L)$ in the equation of a variable y_t^n other than real GDP (y_t^1) are independent $\mathcal{N}\left(0, \tilde{\sigma}_{y^n}^2/\tilde{\sigma}_{y^1}^2\right)$. (The notation $\mathcal{N}(\mu, \sigma^2)$ means a normal distribution with mean μ and variance σ^2.) The prior mean of zero is a neutral benchmark. The variance is analogous to the variance of the Minnesota prior of Litterman (1986): the ratio $\tilde{\sigma}_{y_t^n}^2/\tilde{\sigma}_{y^1}^2$ accounts for the different volatilities of the left-hand-side variable y_t^n and the right-hand-side variable g_t (which is a component of real GDP). Notice that the implicit 'tightness' parameter is equal to 1, so the prior is rather loose: for a variable that is equally volatile as real GDP both the elasticity of 1 and -1 are likely outcomes according to this prior.

The coefficients $a(L)$ in the Phillips curve equation are set as follows. The coefficient of g_{t-1} is $\mathcal{N}(0, \tilde{\sigma}_\pi^2/\tilde{\sigma}_{y_1}^2)$, analogously to the coefficients $b^n(L)$. For the coefficients of g_t and g_{t+1} the prior is tightly concentrated around zero (when we relax their prior, the posterior is concentrated near zero anyway and the marginal likelihood falls).

The prior for the level shift parameter c is $\mathcal{N}(0, 0.5^2)$, i.e. we expect the means of π_t and π_t^e not to differ on average, but this prior is rather loose, with a standard deviation of 0.5 percentage point.

Coefficients of the state equations. In the baseline version of the model the trend of real activity variable y_t^n is a random walk with drift, $\Delta w_t^n = d^n + \eta_t^n$. The drift d^n is $\mathcal{N}(\tilde{\delta}_{y_n}, \tilde{\sigma}_{y_n}^2)$ when y_t^n might be drifting a priori (this is the case for real GDP, investment, imports and exports) and it is fixed at 0 when y_t^n is stationary a priori (unemployment, consumer confidence, capacity utilization).

Trend inflation z_t follows an AR(1) process. The prior for the first order autoregressive parameter f is $\mathcal{N}(0.8, 0.5^2)$. A degree of persistence of 0.8 is a compromise between our prior intuition that trend inflation is very persistent (based e.g. on Cogley et al. 2010) and the persistence of about 0.6 that we find in the training sample. The standard deviation 0.5 includes both quickly mean-reverting and explosive processes. The prior for the constant term d^z is $\mathcal{N}(0.4, 0.5)$. We choose the prior mean of 0.4 because then, when f and d^z are both at their prior means, 0.8 and 0.4 respectively, the implied autoregressive process is stationary with the steady state 2, consistent with the ECB definition of price stability.

The prior about the parameters of the output gap process approximates the ideas from the literature about the periodicity and persistence of the euro area business cycles. The prior is

$$p\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} = \mathcal{N}\left(\begin{pmatrix}1.352\\ -0.508\end{pmatrix}, \begin{pmatrix}0.0806 & -0.0578\\ -0.0597 & 0.0464\end{pmatrix}\right). \tag{B.1}$$

To arrive at this prior we start with the auxiliary model

$$g_t = 2a\cos(2\pi/\tau)g_{t-1} - a^2 g_{t-2} + u_t, \quad u_t \text{ i.i.d. } \mathcal{N}(0,1),\ a > 0,\ \tau > 0. \tag{B.2}$$

This model displays decaying cycles, τ is the periodicity, in quarters, and a is the persistence (the modulus of the root). Harvey et al. (2007) and Planas et al. (2008) advocate the use of this and related parameterizations, because such parameterizations allow specifying priors directly about periodicity and persistence, quantities which are more intuitive than the autoregressive parameters by themselves. Here we follow Planas et al. (2008) and use their prior about $p(\tau, a)$, which is a product of two Beta densities.[12] The prior about τ is centered around 32, implying a business cycle lasting 32 quarters, or 8 years. The prior about a is centered at 0.7. Planas et al. (2008), in turn, base their priors on the analysis of the European output gap performed by Gerlach and Smets (1999) using pre-1998 data. In the second step we arrive at (B.1) by approximating the same dynamics of g using Gaussian priors on ϕ_1, ϕ_2. We find the best approximation following the approach of Jarociński and Marcet (2010).

More in details, let vector g contain the path of the output gap tracked for a specified number of periods T_0. The Planas-Rossi-Fiorentini Beta prior on τ, a implies certain dynamic properties of the output gap, formally summarized by the density $p(g)$. Our goal is to find a Gaussian prior $p(\phi)$ that implies a similar density $p(g)$. Note that we are focusing on approximating $p(g)$, which is what we have priors about, and not on approximating the densities of the parameters of the AR(2) model, which, by themselves, are not inter-

[12]The prior is $(\tau - 2)/(141 - 2) \sim Beta(2.96, 10.70)$ and $a \sim Beta(5.82, 2.45)$, see Planas et al. (2008), p.23.

pretable.[13] Finding the prior for ϕ, $p(\phi)$, means approximating the solution of the integral equation

$$p(g) = \int p(g|\phi)p(\phi)d\phi \tag{B.3}$$

where $p(g|\phi)$, implied by (B.2), is the density of g conditional on a particular value of ϕ. Jarociński and Marcet (2010) propose an efficient iterative numerical procedure for approximating the solution of (B.3) with a density from the desired family, which we choose here to be Gaussian. The outcome of their procedure is the prior (B.1).

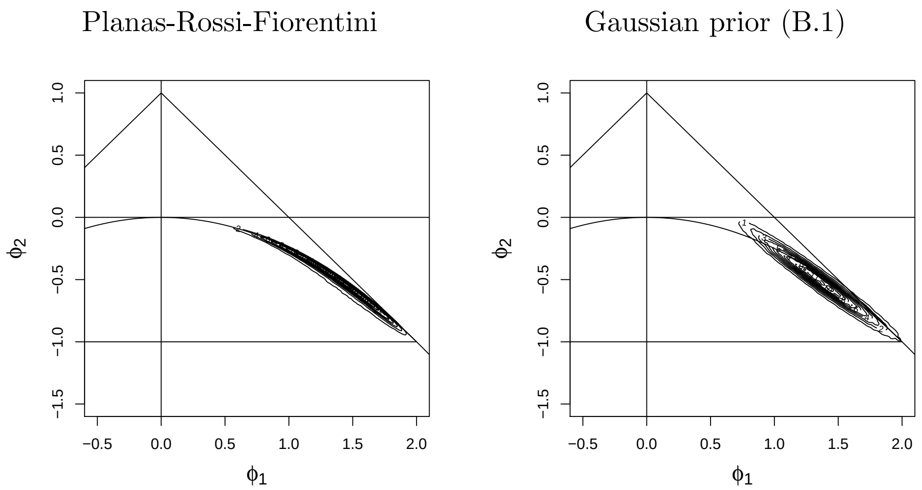
Figure B.1 illustrates the quality of the approximation. Panel A compares the densities of the coefficients ϕ_1 and ϕ_2 implied by (B.2) with the Planas-Rossi-Fiorentini prior (left plot) and Gaussian prior (B.1) (right plot). The Gaussian prior has 0.24 probability mass above the parabola $\phi_1^2 + 4\phi_2 = 0$, i.e. 0.24 probability that the g does not exhibit sinusoidal cycles, while the Planas-Rossi-Fiorentini places probability 1 on such cycles. This might give impression that the Gaussian approximation is poor, but panel B qualifies this impression. Panel B compares the densities of the impulse response, i.e. the dynamics triggered by a unit shock. We can see that the impulse responses look quite similar. We conclude from Panel B that the Gaussian prior (B.2) approximates our prior ideas reasonably well.

Shock variances. When setting the priors about the variances of the shocks we use the rule of thumb that for each observable series v_t, when all the coefficients are at their prior means, the trend and non-trend components account a priori for a half of the variance of Δv_t each, and the variance of Δv_t equals the training sample variance $\tilde{\sigma}_v^2$. We refer to the variance of Δv_t and not of v_t since the series may be non-stationary. All the variances have inverted gamma priors with 5 degrees of freedom, so it remains to specify prior means in order to determine the priors uniquely.[14]

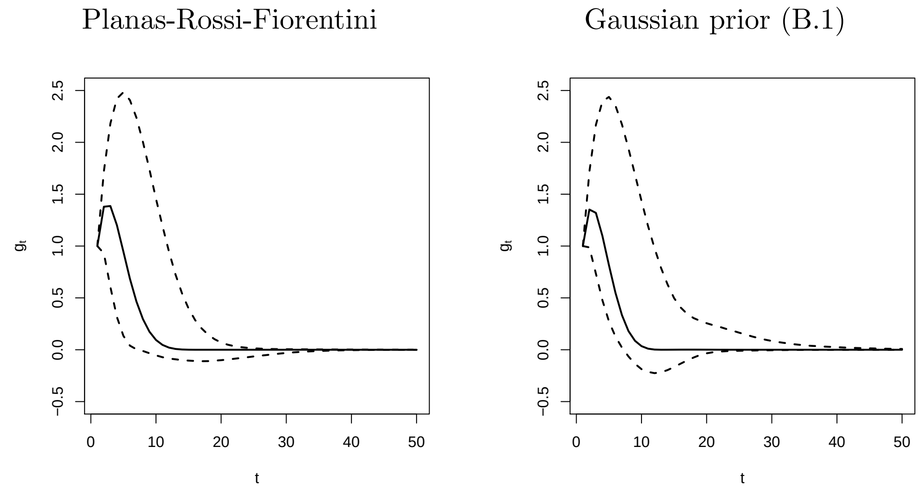
For all variables y^n, $n > 1$ (i.e., other than real GDP), the variances of the shocks in the

[13]To see how important it is to think in terms of the behavior of the modeled variable and not in terms of model parameters, think of the following illustrative example. Consider a process x_t and a model $x_t = \rho x_{t-1} + \varepsilon_t$. Suppose one's prior on the half-life of x_t is centered at 69 periods, corresponding to $\rho = 0.99$. When one thinks of similar models in terms of parameters, one might naively come up with a range $\rho \in (0.97, 1.01)$, as both ends of this range are equally close to 0.99. But values of $\rho \geq 1$ imply infinite half-life. By contrast, when one thinks of similar models in terms of half-life, the range of half-life 69 ± 46 periods corresponds to $\rho \in (0.97, 0.994)$, i.e. a very different range for ρ. This shows that when specifying priors it is important to think in terms of the behavior of the modeled variable and not in terms of model parameters.

[14]For a random variable σ^2 that follows an inverted gamma distribution with scale s and degrees of freedom ν we have that $E(\sigma^2) = s/(v-2)$, so given ν and $E(\sigma^2)$ we set $s = E(\sigma^2)(v-2)$.



A. Joint densities of ϕ_1, ϕ_2. The triangle delimits the stationarity region and the parabola delimits the region of cyclical behavior (see e.g. Hamilton (1994) p.17).



B. Impulse response to a unit shock, median, 10th and 90th percentile.

Figure B.1 – Priors about the dynamics of the output gap: the Planas-Rossi-Fiorentini prior and the Gaussian approximation

trend equation η_t^n and in the observation equation for y_t^n, ε_t^n have means respectively $\tilde{\sigma}^2_{y^n}/2$ and $\tilde{\sigma}^2_{y^n}/4$. To see that these means are consistent with our rule of thumb that half of the variance of Δy^n is explained by the trend and half by the transitory shocks, note that at the prior mean $y_t^n = w_t^n + \varepsilon_t^n = d^n + w_{t-1}^n + \eta_t^n + \varepsilon_t^n$. Then $\Delta y_t^n = d^n + \eta_t^n + \varepsilon_t^n - \varepsilon_{t-1}^n$ and

$\text{var}(\Delta y_t^n) = \text{var}(\eta_t^n) + 2\,\text{var}(\varepsilon_t^n)$. Following the same rule of thumb we set the prior means of the variances of ε^π and ε_t^e to $\tilde{\sigma}_{\pi^e}^2/4$. The prior mean of the variance of the shocks to trend inflation is $\tilde{\sigma}_{\pi^e}^2/2$.

The prior mean of the variance of η_t^g is $0.2\tilde{\sigma}_{y^1}^2$. This mean is consistent with the prior that, conditional on the prior means of ϕ_1 and ϕ_2, g_t accounts for half of the variance of Δy_t^1. To see this, note first that $\text{var}(\Delta y_t^1) = \text{var}(\eta_t^n) + \text{var}(\Delta g_t)$ and $\text{var}(\Delta g_t) = \chi\,\text{var}(\eta_t^g)$ where χ is a function of ϕ_1 and ϕ_2. It is straightforward, though tedious, to show that $\chi = 2(1-\phi_1-\phi_2)/\left((1+\phi_2)((1-\phi_2)^2-\phi_1^2\right)+1$. See e.g. Hamilton (1994) pp.57-58 for similar derivations. Hence, if we want $\text{var}(\Delta g_t) = \chi\,\text{var}(\eta_t^g) = 0.5\tilde{\sigma}_{y^1}^2$ we need to set $\text{var}(\eta_t^g) = 0.5/\chi\tilde{\sigma}_{y^1}^2$ and $0.5/\chi$ evaluates to about 0.2 when $\phi_1 = 1.352$ and $\phi_2 = -0.508$.

Initial states. The prior about the initial states is Gaussian. Let 1 be the first period of the estimation sample. We center the prior for g_1, g_0 and g_{-1} at 0, the prior for w_1 at y_0, and the prior for z_1 at π_0^e. The standard deviations are set to $5\tilde{\sigma}_v$ where v is the respective observable variable. We multiply the standard deviations by 5 in order to make the prior rather diffuse.

Priors in Models 5, 6 and 7. In Model 5, for a variable y^n that has the i.i.d. trend, the prior mean and variance of d^n are equal to the mean and variance of y^n (and not its first difference) in the training sample. In Models 6 and 7, the initial value δ_1^n is centered at $\tilde{\delta}_{y^n}$ with the standard deviation $5\tilde{\sigma}_{y^n}$. The prior means of the variances of ζ_t^n are set to 1000^{-1}, which is a small value. This value implies that it takes on average 1000 quarters for the growth rate of a variable to change by one percentage point. Behind the choice of the variances of ζ_t^n is a separate forecast evaluation exercise, in which we tried the values coded with letters as follows a: 100^{-1}, b: 500^{-1}, c: 1000^{-1}, and d: $10,000^{-1}$. Figure B.2 reports the output gaps estimated on the final sample with each variant of Model 7. As expected, smaller variances of ζ_t^n imply more rigid trends and hence larger output gaps.

We evaluate the out-of-sample forecasts in the same way as described in the main body of the paper. Version c tends to produce the lowest MSEs at longer horizons, and it is also nontrivially different from Model 4, and therefore we pick it to be reported in the main body of the paper labeled simply as ‘Model 7’.

Turning to Model 6, we find that different choices of the variances of ζ_t^n yield very similar

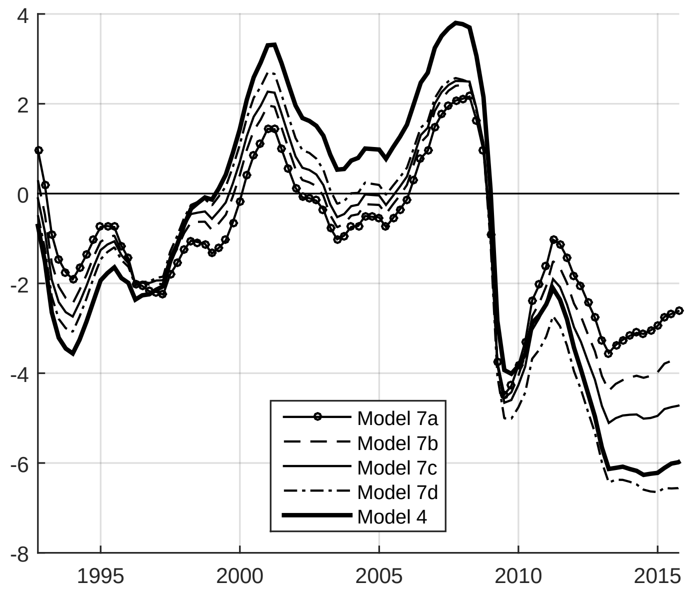


Figure B.2 – Posterior medians of the output gap from different versions of Model 7

output gap estimates, as shown in Figure B.3. For consistency with Model 7, we also pick version c to be reported in the main body of the paper as 'Model 6'.

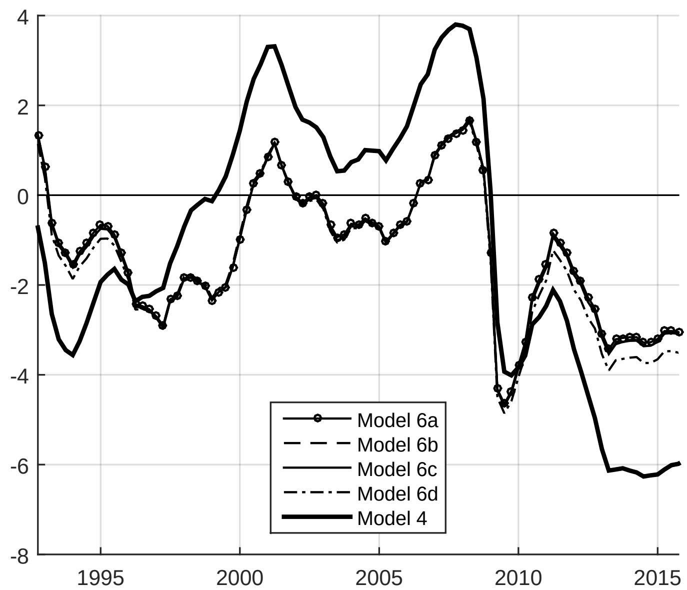


Figure B.3 – Posterior medians of the output gap from different versions of Model 6

Appendix C The posterior distribution of Model 4

Table C.1 reports, following Planas et al. (2008), the posteriors of the parameters of Model 4. We generate 250,000 draws with the Gibbs sampler. We discard the first 50,000 draws, and of the remaining 200,000 we keep every 20th, which leaves us with 10,000 draws. Based on these 10,000 draws, Table C.1 reports for each parameter of Model 4 the mean, the standard deviation, autocorrelations of the draws of order 1 and 50, the relative numerical efficiency (RNE), and Geweke (1992) convergence diagnostics (Z statistics and their p-values). RNE is the ratio of the standard deviation of the mean computed assuming i.i.d. draws to the standard deviation of the mean that takes into account the autocorrelation of the draws. We compute the latter using Newey-West weights with up to 400 lags. Z is the absolute value of the asymptotically normal statistic testing whether the mean based on the first 20% of draws is significantly different from the mean of the last 50% of draws, and its p-value is given in the last column. We never reject at 5% level so chain convergence is obtained. At the bottom of Table C.1 we report the same statistics for the output gap g_T and one-year-ahead inflation forecast $\pi_{T,T+4}$ at the end of the last vintage (T=2015Q4). The contents of the table, the simulation settings and the convergence diagnostics follow closely Planas et al. (2008), except that we have more autocorrelated draws and so we report higher order autocorrelations.

Figure C.1 reports the priors and posteriors of the coefficients of g in the observation equations of real activity variables (coefficients denoted $b^n(L)$). Note that the priors are all weakly informative and centered at zero, while the posteriors are more concentrated and often far away from zero. These posteriors reflect whether a variable is leading, contemporaneous with or lagging the output gap. Consumer confidence is the only leading indicator, with the coefficient on g_{t+1} that is clearly away from zero. The dynamic relationship of Consumer confidence with the cycle is complicated, as it has non-zero loadings also on g_t and g_{t-1}. Real GDP components, investment, imports and exports are clearly contemporaneous with the output gap, as they only have a nonzero coefficient on g_t. The unemployment rate and the capacity utilization contain both contemporaneous and lagging information.

Table C.1 – Posteriors and convergence diagnostics

Parameter	Mean	Sd.	$\rho 1$	$\rho 50$	RNE	Z	$p(z > Z)$
b21	-0.319	0.288	0.26	0.00	0.44	0.21	0.84
b22	2.200	0.425	0.04	0.00	0.82	0.58	0.56
b23	0.202	0.289	0.26	0.00	0.45	0.28	0.78
b31	0.040	0.319	0.29	0.02	0.36	0.75	0.46
b32	3.147	0.491	0.12	0.00	0.55	0.70	0.48
b33	-0.192	0.351	0.38	0.00	0.43	0.39	0.70
b41	0.229	0.394	0.32	0.03	0.35	0.04	0.97
b42	3.295	0.572	0.08	0.01	0.58	0.12	0.90
b43	-0.618	0.410	0.35	0.01	0.47	1.12	0.26
b51	0.013	0.041	0.46	0.04	0.27	1.73	0.08
b52	-0.158	0.054	0.09	0.00	0.67	1.27	0.20
b53	-0.209	0.040	0.45	0.02	0.28	1.13	0.26
b61	2.776	0.767	0.48	0.07	0.26	0.86	0.39
b62	1.899	1.010	0.10	0.01	1.03	0.54	0.59
b63	-1.502	0.771	0.48	0.00	0.32	0.72	0.47
b71	-0.274	0.229	0.55	0.08	0.23	1.10	0.27
b72	1.267	0.307	0.23	0.02	0.59	0.71	0.48
b73	0.765	0.238	0.59	0.04	0.27	1.48	0.14
log var ε^2	-2.011	0.338	0.06	-0.01	0.87	0.91	0.36
log var ε^3	-1.905	0.309	0.01	-0.02	1.20	0.41	0.68
log var ε^4	-1.547	0.279	0.01	0.01	0.87	0.40	0.69
log var ε^5	-6.717	0.321	0.01	0.00	0.99	1.06	0.29
log var ε^6	-0.940	0.406	0.08	0.01	0.94	0.66	0.51
log var ε^7	-3.710	0.415	0.07	-0.01	0.97	0.85	0.40
c	0.214	0.103	0.95	0.13	0.17	0.89	0.38
log var ε^e	-6.117	0.400	0.07	-0.02	0.92	1.12	0.26
ϕ_1	1.735	0.067	0.05	-0.01	0.68	0.39	0.69
ϕ_2	-0.758	0.067	0.04	-0.01	0.74	0.01	0.99
d^1	0.415	0.027	0.07	0.01	0.41	0.22	0.83
d^2	0.382	0.077	0.06	0.01	0.58	0.07	0.94
d^3	1.256	0.083	0.08	0.02	0.50	1.19	0.23
d^4	1.355	0.105	0.06	0.02	0.66	1.91	0.06
d^5	0.000	0.000	-0.02	0.01	1.01	0.40	0.69
d^6	0.000	0.002	0.01	0.01	1.07	1.72	0.08
d^7	0.000	0.001	-0.01	0.01	1.12	0.71	0.48
d^z	0.120	0.037	0.00	0.00	0.84	0.83	0.41
f	0.928	0.019	0.01	0.00	1.00	0.40	0.69
log var η^g	-2.087	0.204	0.13	0.05	0.33	0.42	0.67
log var η^1	-3.141	0.220	0.12	0.02	0.39	0.76	0.45
log var η^2	-0.818	0.241	0.06	-0.02	0.66	1.58	0.11
log var η^3	-0.838	0.258	0.06	0.00	0.62	0.09	0.92
log var η^4	-0.258	0.218	0.02	-0.01	0.90	0.50	0.62
log var η^5	-4.236	0.181	0.06	0.01	0.61	0.70	0.49
log var η^6	1.620	0.170	0.04	0.00	0.95	1.16	0.25
log var η^7	-0.695	0.172	0.04	0.00	0.55	0.14	0.89
log var η^z	-5.473	0.263	-0.01	0.01	0.94	1.09	0.28
az	1.000	0.004	0.01	-0.02	1.19	0.10	0.92
ag1	-0.006	0.010	-0.02	0.00	0.96	1.59	0.11
ag2	-0.003	0.010	-0.02	0.00	1.05	0.86	0.39
ag3	0.123	0.024	0.07	0.04	0.40	0.52	0.60
log var ε^π	-1.401	0.152	0.02	0.00	0.84	0.40	0.69
g_T	-5.951	1.495	0.46	0.08	0.20	0.64	0.52
$100(p_{T+4} - p_T)$	1.025	0.296	0.00	0.02	0.78	0.29	0.77

Note: Statistics based on 10,000 stored draws from the Gibbs sampler. Mean: mean of the draws; Sd.: standard deviation of the draws; ρj: draws autocorrelation of order j; RNE: relative numerical efficiency; Z: Geweke's convergence statistic (absolute value); $p(z > Z)$: p-value of Geweke's convergence statistic, two times the normal cdf evaluated at $-Z$. Coefficients b: the first digit is the number of the real activity variable, the second digit means 1 - lead, 2 - contemporaneous, 3 - lag. The last two entries in the table report selected values of unobservable variables: g_T is the output gap in the last period of the sample and $100(p_{T+4} - p_T)$ is the forecast of inflation 4 periods after the end of the sample.

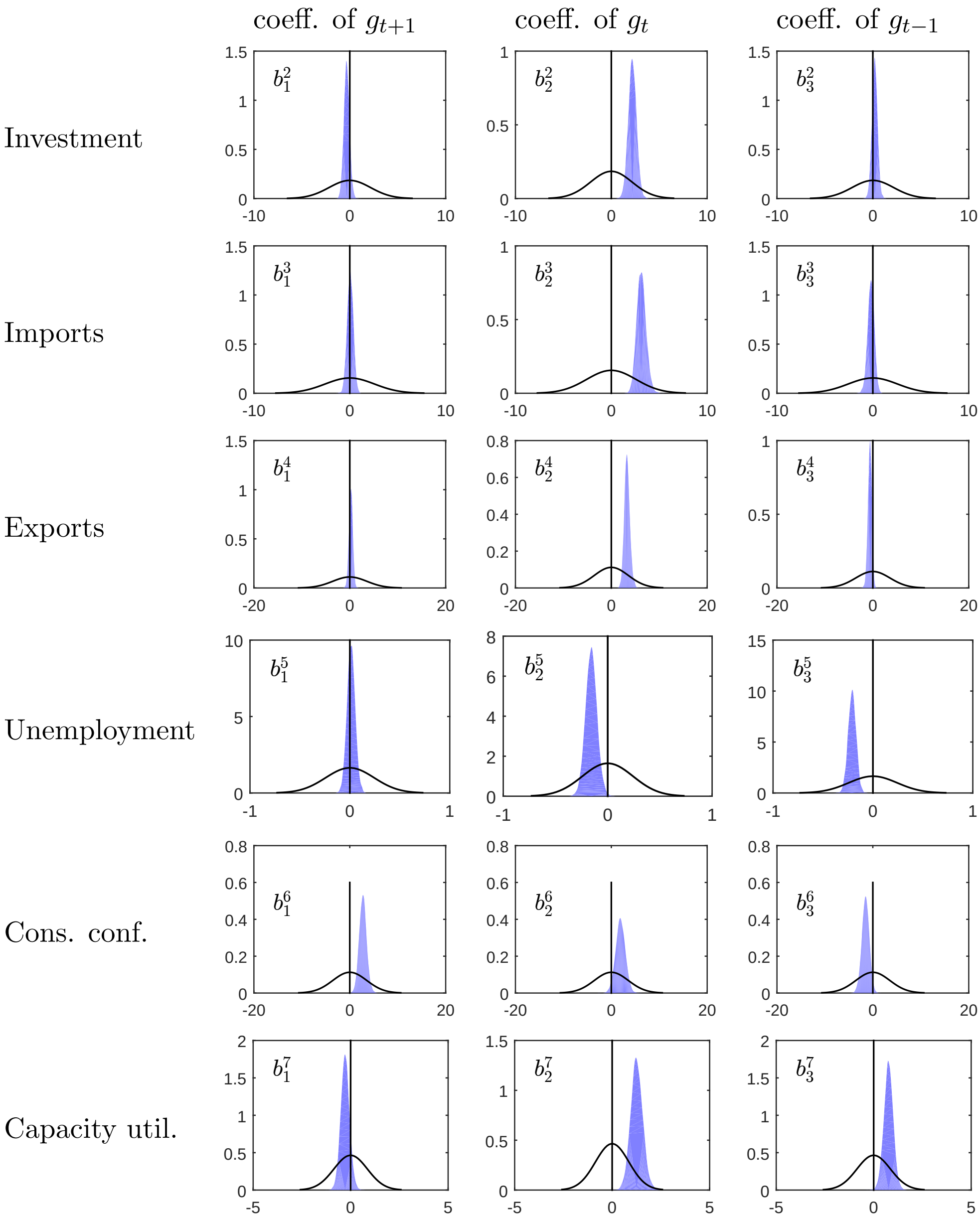


Figure C.1 – Prior and posterior densities of the coefficients of g in the observation equations of real activity variables.

Appendix D The effects of including cost-push variables

In this appendix we report the effects of including cost-push variables in the observation equation for core inflation (1b). We try including the commodity price index, the price of oil and the exchange rate (euros per US dollar). We find that none of these variables is strongly related to core inflation and including them has little effect on the estimates of the output gap.

Figure D.1 plots the cost-push variables and the core inflation. This figure shows that it is not obvious that core inflation is systematically positively related to any of the three standard proxies of cost-push shocks. By eyeballing, we can find both examples and counterexamples of a positive relation.

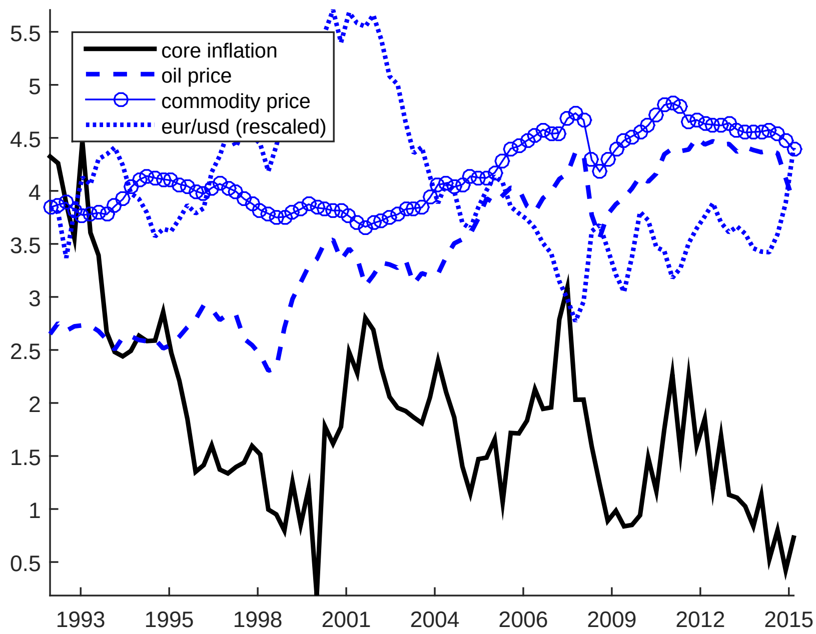


Figure D.1 – Core inflation and three standard proxies for cost-push shocks (in logs)

To study the effect of cost-push variables econometrically we estimate versions of Model 4 with equation (1b) extended to

$$\pi_t = a(L)\,g_t + z_t + a^c(L)\,c_t + \varepsilon_t^{\pi}, \tag{D.1}$$

where c_t is a cost-push variable and $a^c(L)$ is a lag polynomial. We try including the cost-push variables either in log levels (as plotted in the above figure) or in log differences. A priori we think we might need relatively many lags, because e.g. the effect of oil prices on non-energy prices in the core inflation might only materialize with a considerable delay. We report results from the model in differences with 8 lags and from the model in levels with 4 lags, but the results for shorter lag lengths are similar. To be able to track the response of π_t to a shock in c_t we introduce c_t into the system as an endogenous variable, modeling it as an AR(2) with a constant term. We settle for two lags because in most cases the marginal likelihood peaks at two lags.

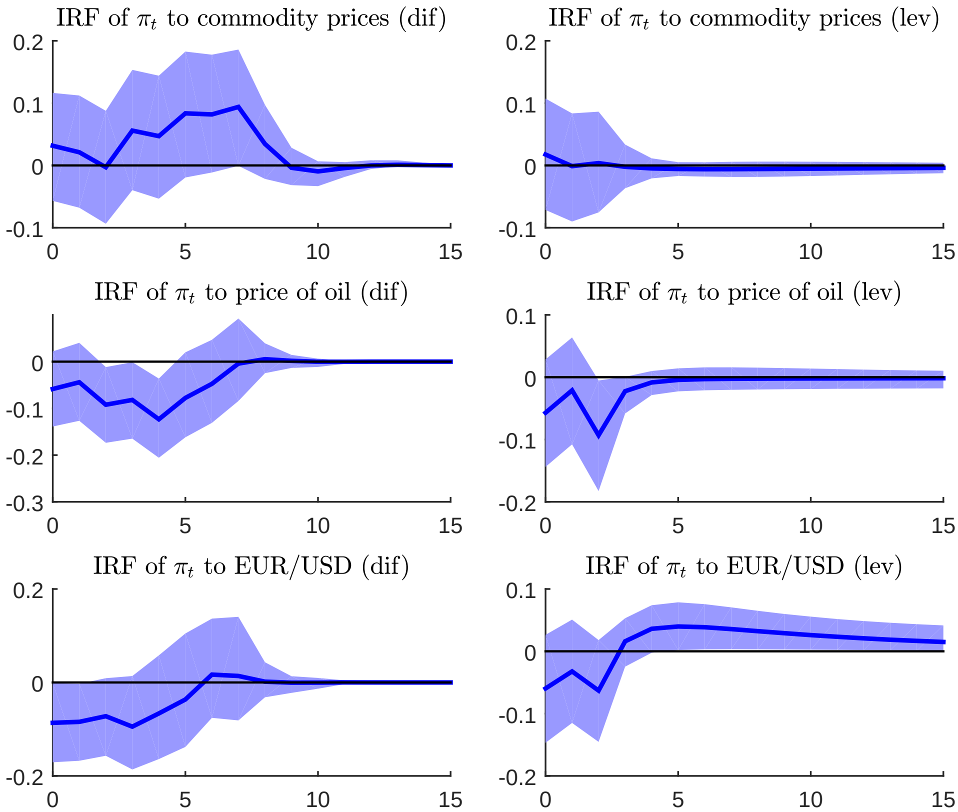


Figure D.2 – Impulse responses of π_t (in %, annualized) to a one standard deviation positive shock in the cost-push variables. 90% posterior probability range. The horizontal axis shows time from the shock, in quarters.

Figure D.2 illustrates the relation of the variables with core inflation. It plots the response of inflation to a one standard deviation positive shock in the cost-push variable. One should expect a positive response of inflation in each case. It turns out that the response of inflation is positive only after a shock to commodity price index when the latter enters in differences, and this response is neither large nor very significant anyway. In the remaining cases the responses are either insignificant or have the wrong sign.

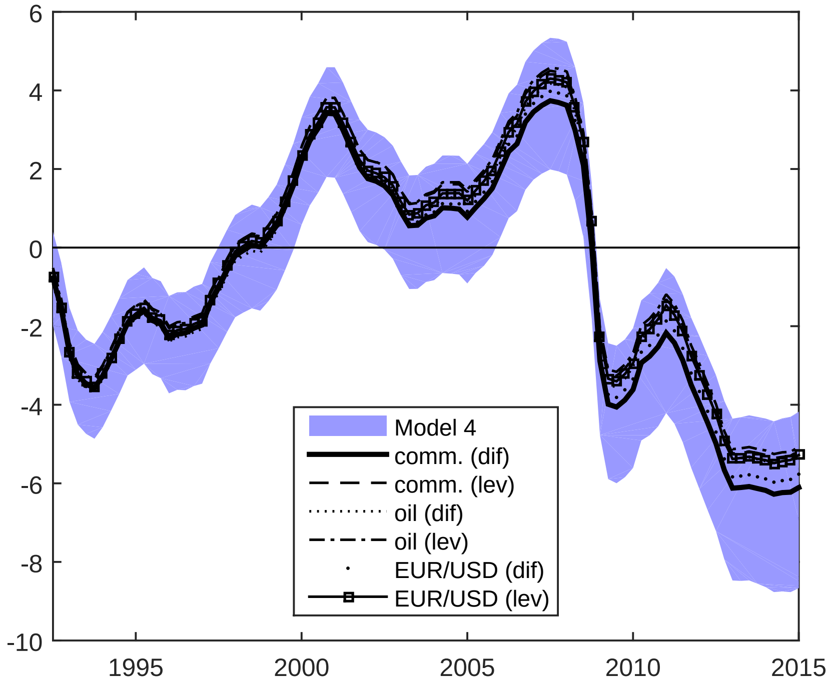


Figure D.3 – Output gap: posterior medians of models with cost-push variables and the 90% posterior probability range from Model 4 (without cost-push variables).

Figure D.3 shows that the estimates of the output gap are not much affected by the inclusion of the cost-push variables. The figure plots the median estimates of the output gap over time, based on the last vintage of the data. We can see that all these estimates are within the 90% band from Model 4, plotted as the blue shaded region. Moreover, the model with commodity prices in differences (black solid line), which is the only one where the sign of the effect of the cost-push variable is correct, is also the closest to the baseline model (we can see that it is closest to the center of the band). To conclude, we find that cost-push variables do not seem to contribute much to our baseline model of core inflation and therefore we omit them in the rest of the paper.

Appendix E Examining the presence of stochastic volatility in our measure of inflation

We estimate the unobserved components - stochastic volatility (UCSV) model of inflation introduced by Stock and Watson (2007). In this model inflation is a sum of a trend τ_t and a transitory component, and the volatility of the shocks to both components varies over time. The model is defined by the following equations,

$$\pi_t = \tau_t + e^{\frac{1}{2}h_t}\varepsilon_t^{\pi}, \tag{E.1a}$$

$$\tau_t = \tau_{t-1} + e^{\frac{1}{2}g_t}\varepsilon_t^{\tau}, \tag{E.1b}$$

$$h_t = h_{t-1} + \omega_h \varepsilon_t^{h}, \tag{E.1c}$$

$$g_t = g_{t-1} + \omega_g \varepsilon_t^{g}, \tag{E.1d}$$

where $\varepsilon_t^{\pi}, \varepsilon_t^{\tau}, \varepsilon_t^{h}$ and ε_t^{g} are independent $\mathcal{N}(0,1)$, and ω_h and ω_g are parameters to be estimated. Following Chan (2016) we estimate this model and evaluate Bayes factors comparing the above model with its versions where the volatilities of either the trend or the transitory component are constant (versions that restrict h_t or g_t to be constant). See Chan's paper for the details on the priors and implementation.[15] Below we report the results for the latest vintage of our inflation variable in the longest estimation sample, 1992Q3-2015Q3, but the conclusions from other vintages are very similar. Table E.1 reports the Bayes factors. This table shows that the full UCSV model (with time varying volatilities in both the trend and transitory component) is basically as good as the model where the variance of the transitory component is fixed (log Bayes factor of -0.2). The full UCSV model is only slightly better than the model in which the variance of the trend component is fixed (log Bayes factor of 1.4). Finally, the full UCSV model is only slightly better than the model with constant variances. The Bayes factor of 1.1 is positive but very small by the standards of Bayes factors. In the Kass and Raftery (1995) classification a log Bayes factor of 1.1 (the Bayes factor of 3) is 'not worth more than a bare mention'.

Figure E.1 shows the estimated evolution of the volatilities. There is not much meaning-

[15] We used the Matlab programs accompanying Chan (2016), downloaded from http://people.anu.edu.au/joshua.chan/.

Table E.1 – Log Bayes factors in favor of model (E.1a-E.1d) compared with its restricted versions.

Model in (E.1a-E.1d) vs ...	h_t constant	g_t constant	h_t and g_t constant
log BF	-0.2	1.4	1.1
numerical standard error	(0.01)	(0.02)	(0.04)

ful variation in the volatility of the transitory component $\exp(h_t/2)$. The volatility of the trend component $\exp(g_t/2)$ is larger in the early 1990s, when trend inflation was falling, and by 1995 falls to a lower level and remains approximately constant through the rest of the sample. For headline inflation measures researchers often find an increase of the volatility during the Great Recession, but we can see that this finding does not extend to our measure of core inflation.

We have repeated the same exercise for headline inflation and find much larger changes in volatility and much higher Bayes factors in favor of stochastic volatility. We have also repeated the exercise for the longest possible sample starting in 1985, combining our training sample and the estimation sample, and find somewhat higher Bayes factors in favor of stochastic volatility.

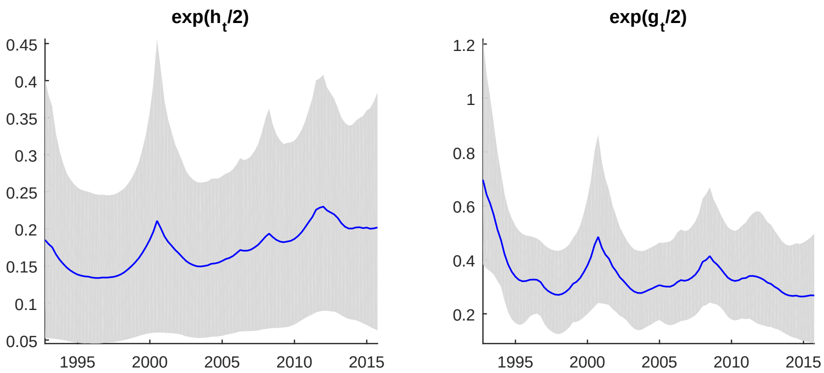


Figure E.1 – The estimates of $\exp(h_t/2)$ (left panel) and $\exp(g_t/2)$ (right panel). Median and 90% posterior probability range.

Appendix F Forecasting the whole set of variables: marginal likelihood

While Model 4 generates the best real-time inflation forecasts, Model 7 attains a higher marginal likelihood. Table F.1 reports the marginal likelihoods of Models 4, 5, 6 and 7 computed on the last vintage of the data. (Recall that marginal likelihoods are comparable only across models that have the same observables, hence we have to exclude Models 1, 2 and 3, which have different observables.) Marginal likelihood can be written as a product of one-step-ahead out-of-sample predictive densities of all the observables in the model. Hence, the fact that Model 7 attains a higher marginal likelihood than Model 4 suggests that relaxing the assumption of constant drifts in the trend processes improves the predictive densities of some other variables, while sacrificing some of the real-time predictive performance for inflation.

Table F.1 – Marginal likelihood of Models 4, 5, 6 and 7, last vintage.

Model number, trend type	log marginal likelihood
Model 4: RW	-892
Model 5: RW or i.i.d	-1082
Model 6: IRW	-923
Model 7: LLT	-867

Note: The marginal likelihoods of Models 4-7 can be compared because these models have the same observables. We do not present the marginal likelihoods of Models 1, 2 and 3 as these models have different observables and hence are not comparable.

References

Basistha, A. and Startz, R. (2008). Measuring the nairu with reduced uncertainty: A multiple-indicator common-cycle approach. *The Review of Economics and Statistics*, 90(4):805–811.

Baştürk, N., Çakmakli, C., Ceyhan, S. P., and Van Dijk, H. K. (2014). Posterior-predictive evidence on US inflation using extended New Keynesian Phillips curve models with non-filtered data. *Journal of Applied Econometrics*, 29(7):1164–1182.

Blanchard, O. (2015). Ten takeaways from the 'Rethinking macro policy. Progress or confusion?'. Voxeu article, International Monetary Fund.

Chan, J. C., Clark, T., and Koop, G. (2015). A new model of inflation, trend inflation, and long-run inflation expectations. Working paper no. 15-20, Federal Reserve Bank of Cleveland.

Chan, J. C. C. (2016). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*. forthcoming.

Clark, T. E. and Doh, T. (2014). Evaluating alternative models of trend inflation. *International Journal of Forecasting*, 30(3):426–448.

Cogley, T., Primiceri, G. E., and Sargent, T. J. (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, 2(1):43–69.

D'Agostino, A., Giannone, D., Lenza, M., and Modugno, M. (2015). Nowcasting business cycles: a Bayesian approach to dynamic heterogeneous factor models. *Advances in Econometrics*. forthcoming.

Diron, M. and Mojon, B. (2008). Are inflation targets good inflation forecasts? *Economic Perspectives, Federal Reserve Bank of Chicago*, 32(2Q):33–45.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–615.

Fagan, G., Henry, J., and Mestre, R. (2001). An area-wide model (AWM) for the euro area. Working Paper Series 0042, European Central Bank.

Faust, J. and Wright, J. H. (2013). Chapter 1 - forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, Part A of *Handbook of Economic Forecasting*, pages 2 – 56. Elsevier.

Fischer, B., Lenza, M., Pill, H., and Reichlin, L. (2009). Monetary analysis and monetary policy in the euro area 1999-2006. *Journal of International Money and Finance*, 28(7):1138–1164.

Gerlach, S. and Smets, F. (1999). Output gaps and monetary policy in the EMU area. *European Economic Review*, 43(4-6):801–812.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Berger, J., Bernardo, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics*, pages 169–193. Oxford University Press. Proceedings of the Fourth Valencia International Meeting on Bayesian Statistic.

Giannone, D., Henry, J., Lalik, M., and Modugno, M. (2012). An area-wide real-time database for the euro area. *The Review of Economics and Statistics*, 94(4):1000–1013.

Giannone, D., Lenza, M., Momferatou, D., and Onorante, L. (2014). Short-term inflation projections: A Bayesian vector autoregressive approach. *International Journal of Forecasting*, 30(3):635–644.

Gordon, R. J. (2014). The turtle's progress: Secular stagnation meets the headwinds. In Coen, T. and Baldwin, R., editors, *Secular Stagnation: Facts, Causes and Cures*, pages 47–60. VOX CEPRs Policy Portal.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Harvey, A. C., Trimbur, T. M., and Van Dijk, H. K. (2007). Trends and cycles in economic time series: A Bayesian approach. *Journal of Econometrics*, 140(2):618–649.

Jarociński, M. (2015). A note on implementing the Durbin and Koopman simulation smoother. *Computational Statistics & Data Analysis*, 91(C):1–3.

Jarociński, M. and Marcet, A. (2010). Autoregressions in small samples, priors about observables and initial conditions. Working Paper Series 1263, European Central Bank.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kuttner, K. N. (1994). Estimating potential output as a latent variable. *Journal of Business & Economic Statistics*, 12(3):361–68.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business and Economic Statistics*, 4(1):25–38.

Mertens, E. (2014). On the reliability of output gap estimates in real time. Unpublished manuscript, Federal Reserve Board.

Morley, J. C., Nelson, C. R., and Zivot, E. (2003). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *The Review of Economics and Statistics*, 85(2):235–243.

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.

Planas, C., Rossi, A., and Fiorentini, G. (2008). Bayesian analysis of the output gap. *Journal of Business and Economic Statistics*, 26(1):18–32.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335.

Stock, J. H. and Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33.

Summers, L. (2013). Why stagnation might prove to be the new normal. Lawrence Summers' blog, December 15, 2013, Harvard University.

Valle e Azevedo, J., Koopman, S. J., and Rua, A. (2006). Tracking the business cycle of the euro area: A multivariate model-based bandpass filter. *Journal of Business & Economic Statistics*, 24:278–290.

Acknowledgements
For comments and suggestions, we thank Carlo Altavilla, Marta Bańbura, Gabriel Fagan, Domenico Giannone, Siem Jan Koopman, Laurent Ferrara, Bartosz Maćkowiak, Giorgio Primiceri, Lucrezia Reichlin, Mathias Trabandt, an anonymous referee of the ECB working paper series, and participants at several seminars and conferences. The opinions in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank and the Eurosystem.

Marek Jarociński
European Central Bank; email: marek.jarocinski@ecb.europa.eu

Michele Lenza
European Central Bank and Université Libre de Bruxelles, ECARES; email: michele.lenza@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 at 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.

ISSN 1725-2806 (pdf)
ISBN 978-92-899-2214-2 (pdf)
DOI 10.2866/944018 (pdf)
EU catalogue No QB-AR-16-083-EN-N (pdf)