# Output gap and inflation forecasts in a Bayesian dynamic factor model of the euro area*

Marek Jarociński† Michele Lenza‡

Preliminary, this version: July 3, 2015

**Abstract**

We estimate a Bayesian dynamic factor model of the euro area core inflation and real activity. The common cyclical factor is normalized so that it coincides with the deviation of output from its trend, and hence we call it the output gap. We examine the precision of inflation forecasts and the robustness of this output gap in real-time. We find that it helps to introduce multiple activity indicators, to relate trend inflation to long term inflation expectations, and to model trends of the remaining variables as random walks. The resulting model forecasts inflation well and implies that the output gap in the euro area has been as large as -6% in 2014.

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

**Keywords**: output gap, Phillips curve, factor model, inflation forecast

---

*We would like to thank Rafaella Giacomini, Domenico Giannone, Siem Jan Koopman and Giorgio Primiceri for their comments. 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.

†European Central Bank, email:marek.jarocinski@ecb.europa.eu

‡European Central Bank and Université Libre de Bruxelles, ECARES, email:michele.lenza@ecb.europa.eu

# 1 Introduction

Economists monitor and assess real economic activity in a variety of ways and for a variety of important reasons. In this paper we focus on only one of these reasons: the slack in real economic activity is potentially useful for forecasting inflation. This idea follows at least from Phillips (1958), see also e.g. Stock and Watson (1999), and is reflected in countless academic discussions and policy analyses.

We use a small Bayesian factor model to efficiently summarize multidimensional economic activity in a single indicator, the output gap, and to forecast core inflation with it. The model includes a vector of real activity indicators and core inflation. Long run behavior of the variables is captured by variable-specific persistent trends. Fluctuations of the variables at business cycle frequency around their trends are captured by a common factor. Normalization ensures that this common factor coincides with the current deviation of real GDP from its trend, and is therefore our measure of the output gap. The model is flexibly specified to account for the possibility that other variables may be either leading or lagging the real GDP. We apply the model to the euro area econonomy.

We face a number of modeling choices: Which observable variables to include? How to specify the trend processes? What priors to use? We find that different reasonable choices lead to widely disparate estimates of the output gap in the euro area. These different estimates approximately agree about the timing of peaks and troughs, but often widely disagree about the level of the output gap. Given that the output gap is unobservable, how is one to judge which of these estimates is the most useful?

The approach we take in this paper is that an output gap estimate is only as useful as the real-time signal it sends about future inflation. Therefore, we use out-of-sample, real-time forecasts of core inflation as a validation tool for our output gap models. We find that the best output gap estimates are extracted from a relatively large set of observable variables, with relatively inflexible trend processes, and it is useful to relate trend inflation to long term inflation expectations. The resulting forecasts of inflation are very good, both before the 2008 crisis, and since its onset, when we correctly capture the fall in inflation. The output gap in the crisis is large and by 2014 it may be as large as -6% of euro area GDP!

Whether slow growth results from a large output gap or slow trend growth matters fundamentally for economic policy makers. A large output gap calls for a demand stimulus, while slow trend growth probably requires supply-side policies. Many economists 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)). We find that specifications of our model that produce slow trend growth and, consequently, small output gaps, forecast core inflation poorly. Therefore, our results highlight that reconciling the above version of the secular stagnation hypothesis with the core inflation data remains a challenge.

We also study real-time reliability of end-of-sample estimates of the output gap, which is crucial if they are to be of use for policy. In their influential paper, Orphanides and van Norden (2002) demonstrate 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. To confront this issue convincingly we take a fully real-time perspective in our econometric analysis. In particular, we update and use the real-time database for the euro area described in Giannone et al. (2012), which collects the data appearing in the ECB monthly bulletin since 2001. 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 they are 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 modeling 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) our model accounts for the presence of both leading, coincident 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.

The output gap concept in the reduced form approaches like ours and the ones cited above is different from the output gap concepts in structural dynamic stochastic general equilibrium (DSGE) models. A simple way to see this is to compare the New-Keynesian Phillips curve equation present in these models with the traditional Phillips curve present in our model. The New-Kenyesian Phillips curve includes the next-period inflation expectations, such as e.g. in $\hat{\pi}_t = \beta E_t \hat{\pi}_{t+1} + \kappa \tilde{g}_t + u_t$, where $\hat{\pi}_t$ is the deviation of inflation from its trend in period $t$, $\tilde{g}_t$ is the appropriate output gap concept, $u_t$ collects the shocks to this equation, and $\beta$ and $\kappa$ are coefficients. The Phillips curve equation in our model reads $\hat{\pi}_t = \kappa g_t + u_t$ where $g_t$ is our output gap. Clearly, as these are different equations, only by chance would $\tilde{g}_t$ and $g_t$ coincide. The output gap in DSGE models has often been contrasted with reduced form output gaps and indeed their time series properties can in principle be very different. For example, Christiano et al. (2010) show how the DSGE-model consistent output gap can be either procyclical or countercyclical depending on model assumptions. On the other hand, e.g. Galí et al. (2012) and Justiniano et al. (2013) produce output gap measures that look similar to the output gaps from reduced form approaches.

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

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

The first block of our database consists of seven indicators of real economic activity that we collect in the vector $y_t$: real GDP ($y_t^1$), real private investment ($y_t^2$), real imports ($y_t^3$), real export ($y_t^4$), unemployment ($y_t^5$), 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 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 5-year ahead inflation expectations ($\pi_t^e$) for the euro area from Consensus Economics. Consensus Economics collects and publicly releases 5-year inflation forecasts of G-7 countries every April and October since 1989. We compute the 5-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 missing.

For each variable, we collect the 55 real-time data vintages released in the beginning of the third month of each quarter from 2001Q1 to the 2014Q3.[1] Consequently, for the last quarter of each real-time sample we only observe capacity utilization (which is a survey) and inflation expectations (in the second and fourth quarter of the year, otherwise the current 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 observations from 1985Q1 to 1992Q2 are used as a training sample, to inform our prior. Observations starting from 1992Q3 are used for the estimation.

# 3 Econometric model

We use the following state space model to estimate output gap and to forecast inflation. The observation equations of the model are

$$y_t = B^n(L)g_t + w_t^n + u_t^{y^n}, \tag{1}$$

$$\pi_t = a(L)g_t + z_t + u_t^{\pi}, \tag{2}$$

$$\pi_t^e = c + z_t + u_t^e, \tag{3}$$

where $u_t^{y^n}, u_t^{\pi}, u_t^e$ are independent Gaussian errors and $L$ denotes the lag operator. $n$, the index of real activity variables, ranges from 1 to $N$.

The first equation relates the real "n-th" real activity variable $y_t^n$ to a variable

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

specific common trend $w_t^n$ and to $g_t$, a common factor. The latter may enter the equations at all leads and lags and, hence, $B(L)$ is a polynomial with both negative and positive powers of $L$. In so doing, we accommodate for the presence of both contemporaneous, lagging and leading indicators in the vector of real activity variables $y_t$. The first variable, $y_t^1$, is the log of the real GDP and for this variable we restrict the coefficients of $g_t$ to be 1, the coefficients 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 and ensures that it is expressed in percent of real GDP.

The second equation, referred to as the Phillips Curve, relates inflation to the current and lagged output gap $g_t$ and to trend inflation $z_t$. This specification differs from the popular "triangle" model of inflation (Gordon, 1997) that relates inflation to output gap, lagged inflation and cost push variables. In our model, the persistence of inflation is accounted for by the persistence of trend inflation and the persistence of the output gap, so we do not need lagged inflation. Moreover, our measure of inflation is based on HICP excluding energy and food and we found that including cost-push variables, such as the oil price and the exchange rate did not improve our inflation forecasts

The third equation relates trend inflation $z_t$ to long term inflation expectations $\pi_t^e$, as advocated by e.g. Faust and Wright (2013) and Clark and Doh (2014). A theoretical justification for such a model of trend inflation is provided in e.g. Cogley et al. (2010). We could introduce $\pi_t^e$ directly into the Phillips curve, but we prefer to have $\pi_t^e$ among the endogenous variables so that our model can provide a forecast of inflation without having to impose any off-model future path for the inflation expectations. $z_t$ differs from $\pi_t^e$ by a level shift $c$ and i.i.d. noise $u_t^e$, to account for the fact that inflation expectations $\pi_t^e$ and inflation $\pi_t$ refer to different concepts of inflation (headline HICP and HICP excluding energy and food).

The state equations of the model are the following. The output gap is modeled as an autoregressive process of order two,

$$g_t = \phi_1 g_{t-1} + \phi_2 g_{t-2} + u_t^g. \tag{4}$$

Trends of the real activity variables are modeled alternatively as constants, $w_t^n = d^n$, random walks with drift, $\Delta w_t^n = d^n + u_t^{w,n}$ or second order random walks, $\Delta\Delta w_t^n = u_t^{w^n}$. Trend inflation $z_t$ is modeled as an autoregressive process of order one, $z_t = d^z + f z_{t-1} + u_t^z$. All the shocks are Gaussian.

## 3.1 Priors

Overall, our prior selection 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 VARs. Here we only sketch the most relevant aspects, while the full details on the specification are reported in appendix B at the end.

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 a loose variant of the Minnesota prior. 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, which is important in our case, since, as explained below, we recompute the posteriors hundreds of times.

## 3.2 Estimation

In the Gibbs sampler we draw the parameters ($B^n(L)$ with $n = 1, ..., N$, $a(L)$, $c$, $\phi_1$, $\phi_2$, $d$, $f$, and all the shock variances) conditionally on the unobserved states ($g_t$, $w_t^n$ with $n = 1, ..., N$, $z_t$ 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 (2014). To compute each posterior we generate 250,000 draws with this Gibbs sampler, out of which we discard the first 50,000. We confirm the convergence of the Gibbs sampler using the Geweke (1992) diagnostics.

# 4 Empirical results

## 4.1 Model specifications

We estimate six 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 6 feature both long term inflation 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 Model 5 are more rigid, and in Model 6 more flexible. In Model 5 trends of the a priori stationary real activity variables

Table 1: Model specifications.

| | trend $y$ | trend $\pi$ related to | Variables $\pi_t$ | $\pi_t^e$ | $y_t^1$ | $y_t^2$ | $y_t^3$ | $y_t^4$ | $y_t^5$ | $y_t^6$ | $y_t^7$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Model 1 | RW | - | x | | x | | | | | | |
| Model 2 | RW | $\pi_t^e$ | x | x | x | | | | | | |
| Model 3 | RW | - | x | | x | x | x | x | x | x | x |
| Model 4 | RW | $\pi_t^e$ | x | x | x | x | x | x | x | x | x |
| Model 5 | RW or constant | $\pi_t^e$ | x | x | x | x | x | x | x | x | x |
| Model 6 | 2nd order RW | $\pi_t^e$ | 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. $\pi_t$ is the quarterly percentage change in HICP excluding enery and food; $\pi_t^e$ is the five years ahead inflation expectations 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.

(unemployment rate, consumer confidence and capacity utilization) are modeled as constants.[2] In Model 6, the trends of all the real activity variables are modeled as a second order random walk, which is a more flexible process.[3]

## 4.2 Output gap estimates on the last vintage of the data

We start by estimating each of the six models just described on the most recent sample, 1992Q3 to 2014Q3. Figure 1 plots the point estimates (posterior medians) of the output gap over time obtained from each of the six models. 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 of utmost important to discriminate among the model features we have discussed above because different mixes of trend specifications and observables lead to substantial disagreements about the size of the output gap. For example, at the end of the sample, the estimates of the output gap

[2]The prior mean and variance of these constants are equal to the mean and variance of the corresponding variable in the training sample.

[3]The initial value $\Delta\Delta w_1^n$ is centered at $\tilde{\delta}_y$ with standard deviation $5\tilde{\sigma}_y$. The priors for the shock variances in these second order random walks are set to 1/500, which is a small value. This value implies that it takes on average 500 quarters for the growth rate of a variable to change by one percentage point.

range between 0 and -6 percent of GDP.

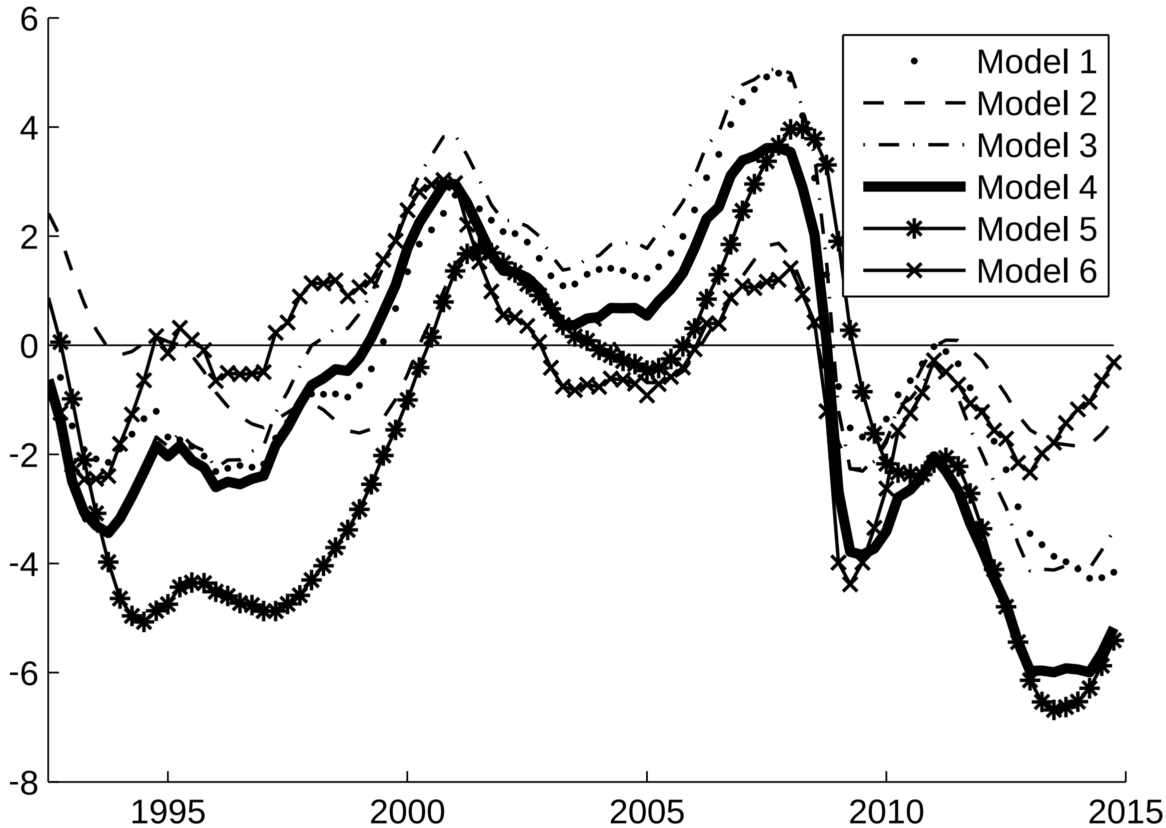



Figure 1: Point estimates (posterior medians) of the output gap from Models 1 to 6

### 4.3 Forecasting results

In this subsection, we study the real-time out-of-sample forecasting performance of our models in order to discriminate among the different measures of output gap. We proceed as follows. We re-estimate each of the six models over 55 expanding samples of our real-time data and, at each point in time, we forecast inflation up to two years ahead. The first estimation sample (data available on 2001Q1) spans the period 1992Q3 - 2001Q1 and the last (data available on 2014Q43) 1992Q3 - 2014Q3.[4] Our

[4]As explained in the data section, however, we have a ragged edge at the end of the sample, due to the different timeliness in the data releases. The only data release for the current quarter is

target measure of inflation for horizon $h$ is the annualized rate of change in consumer prices $\pi_{t,t+h}$ defined as

$$\pi_{t,t+h} = \frac{4}{h}\left(p_{t+h} - p_t\right),$$

where $p_t$ the log-level of consumer prices. We compute the target inflation rate using the last data vintage (2014Q3). We evaluate both the point and the density forecasts.

We start with the evaluation of point forecasts. Table 2 reports the mean squared error (MSE) of the nowcast of inflation ($h = 0$),[5] the four ($h$=4) and eight ($h$=8) quarters ahead forecasts. The point forecasts are computed as the median of the posterior predictive density. The results are cast in terms of ratios of MSE of the different models over the MSE of the simple benchmark forecast. The benchmark forecasts come from the random walk with drift for $p_t$.[6] A number smaller than one indicates that the model outperforms the simple benchmark.

Table 2: MSE relative to the simple benchmark.

| **Model** | $h = 0$ | $h = 4$ | $h = 8$ |
|---|---|---|---|
| Model 1 | 0.62 | 1.16 | 2.66 |
| Model 2 | 0.64 | 1.12 | 1.62 |
| Model 3 | 0.61 | 1.10 | 2.26 |
| Model 4 | 0.52 | 0.56 | 0.73 |
| Model 5 | 0.62 | 0.58 | 0.86 |
| Model 6 | 0.93 | 1.08 | 0.97 |

Note: Ratio of mean squared forecast error of Models 1 to 6 relative to the random walk with drift for $p_t$. Current quarter ($h$=0), four quarters ahead ($h$=4) and eight quarters ahead ($h$=8) forecasts. Numbers smaller than one indicate that the model outperforms the random walk.

Table 2 shows that the Phillips curve inflation forecasts (inflation forecasts using activity variables) can outperform simple benchmarks in the euro area, but the spec-

capacity utilization.

[5]For example, for the vintage of data available in 2001Q1 the nowcast refers to the inflation rate in 2001Q1, i.e. the current quarter. In this vintage only capacity utilization is available for 2001Q1 and the latest observation of inflation is from 2000Q4.

[6]The random walk with drift for $p_t$ is the best of the standard simple benchmarks and produces forecasts that are very difficult to beat in the euro area, see e.g. Diron and Mojon (2008), Fischer et al. (2009), Giannone et al. (2014).

ification of the model matters crucially for the forecasting performance. First, the models including the full set of real economy variables (Models 3 to 6) are generally performing better 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 four and eight quarters 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 6, which do include the expectations. The final lesson we draw from the evaluation of the point forecasts is that allowing for the more flexible trend representations embedded in Model 6, compared to the rest of the models, does not pay in terms of forecasting accuracy. 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 inflation trend and a parsimonious random walk representation for the trends in the real economy variables. Model 4 is our baseline model.

Figure 2 presents the quantiles of the predictive density of inflation four quarters ahead, along with the actual inflation. The solid line shows the annual inflation. The value of the solid line in period $t$ represent $100(p_t - p_{t-4})$. The dashed lines show percentiles 50, 16 and 84 of the predictive density of inflation. The values of the dashed lines in period $t$ represents the density of $100(p_t - p_{t-4})$ generated with the real time data available in $t - 4$. Two main lessons follow from this figure. First, it shows that four quarters ahead forecasts become much less volatile when *two* elements are present: we use multiple indicators of real activity and we relate trend inflation to long term inflation expectations. To see this, note that inflation forecasts are much more volatile in Models 1 to 3 than in Models 4 to 6. Second, the forecasts from

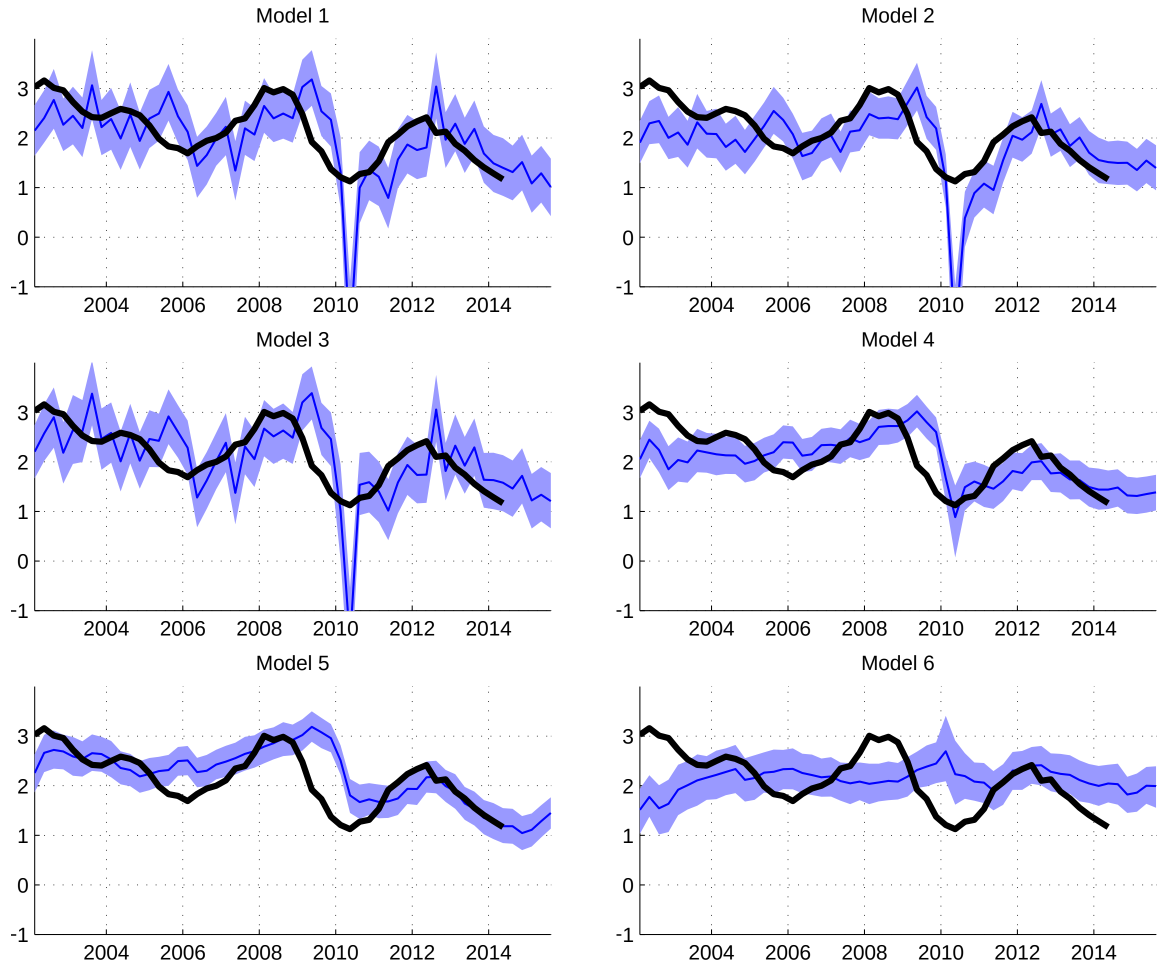



Figure 2: Four quarters ahead forecasts of inflation (blue line: median, blue shaded area: percentile 16 to 84) and actual inflation (black line)

Models 4 and 5 are rather similar and track inflation better than the forecasts from Model 6. However, Model 5 predicts a much too high inflation in 2010, while Model 4 matches that episode almost perfectly.

Table 3 complements the picture by presenting log scores of predictive densities. This table confirms that Model 4 dominates other specifications. At the four quarter horizon Models 5 and 6 lose with Models 1 to 3 according to the log score, even though they produce better point forecasts. The ranking is changed because of the different behavior of predictive variances. The variances of the forecasts from Models

Table 3: Average difference of log-scores compared with the simple benchmark.

| **Model** | $h = 0$ | $h = 4$ | $h = 8$ |
|---|---|---|---|
| Model 1 | 0.31 | 0.26 | 0.22 |
| Model 2 | 0.31 | 0.24 | 0.37 |
| Model 3 | 0.31 | 0.21 | 0.13 |
| Model 4 | 0.38 | 0.29 | 0.35 |
| Model 5 | 0.32 | 0.02 | -0.17 |
| Model 6 | 0.13 | -0.06 | 0.32 |

Note: Average difference of log-scores of Models 1 to 6 from the log-score of the random walk with drift for $p_t$. Current quarter ($h$=0), four quarters ahead ($h$=4) and eight quarters ahead ($h$=8) forecasts. Numbers bigger than zero indicate that the model outperforms the random walk.

5 and 6 are quite low when they make the largest mistake, in 2010. By contrast, the variances of the forecasts from Models 1 and 3 are very large in 2010, which cushions the effect of the forecast error on the log score. Model 4, by contrast, has both low variance and good precision of the point forecast, and hence outperforms the other models in terms of log scores.

Table 4 reports the marginal likelihoods of Models 4, 5 and 6 computed on the last vintage of the data. The marginal likelihood is a function of the out-of-sample predictive density scores for all the variables in the model, hence it is only comparable across models that have the same variables (i.e. Models 4 to 6). The lesson from this table is that, also according to the marginal likelihoods, Model 4 is preferable to Models 5 and 6, i.e. that simple random walk trends are preferable to either more or less restrictive specifications of trends.

Table 4: Marginal likelihood of Models 4, 5 and 6, latest sample.

| Model and main features | log marginal likelihood |
|---|---|
| Model 4: random walk all variables | -860 |
| Model 5: stationary variables constant, rest random walk | -1030 |
| Model 6: second order random walk | -957 |

Note: The marginal likelihoods of Models 4, 5 and 6 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.

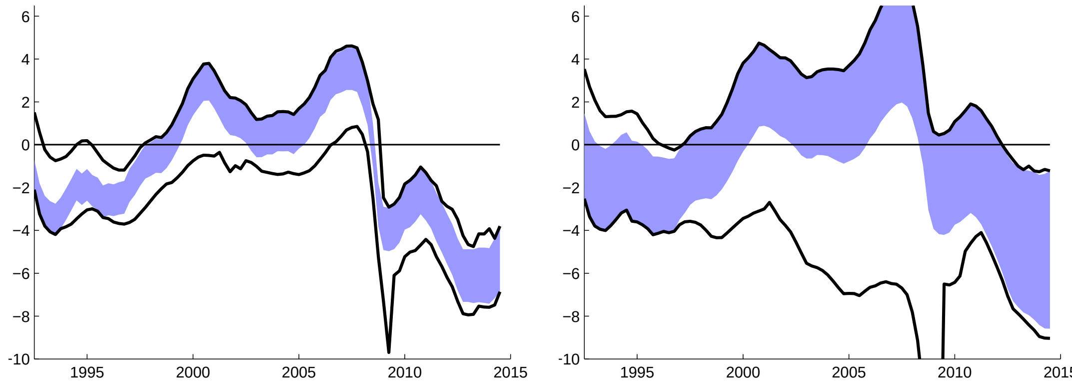
## 4.4 Robustness of the output gap

This subsection reports the robustness of the output gap estimates in real time. The issue of robustness has attracted much attention especially 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. Therefore, we study the robustness of our estimates of output gap in real time. We find that in some models robustness is indeed a serious concern. However, output gap estimates from our best performing model, Model 4, turn out to be quite robust in real time.

To summarize the real time revisions of the output gap we compute the envelope of the 16th and 84th percentiles coming from our 55 real time samples. We compute the envelope as follows. We have 55 posteriors and hence 55 sets of posterior quantiles of the output gap, obtained with 55 real time samples. At each date, we take the lowest of the available 16th percentiles in that date, 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 (199Q3-2014Q3). 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, as shown in the left panel, the lessons about the output gap coming from Model 4 are robust in real time.

## 4.5 The output gap in the current crisis: what do we learn

According to our best performing model, Model 4, the output gap is large at the end of the sample. Taken at face value, this finding suggests that presently a demand stimulus that would close this output gap is more urgent than structural reforms. This view is, however, not uncontroversial. Many analysts and policy makers believe that a crucial problem facing the euro area is that trend GDP growth has stalled and




Model 4 Model 1

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)

structural reforms are needed to revive it. Figure 4 illustrates these alternative views. Model 6 is consistent with the view that trend GDP growth has stalled. This view is also intuitive, but we find that it produces worse inflation forecasts.

# 5 Conclusions

We estimate the output gap in the euro area with several specifications of a small Bayesian dynamic factor model. We find that while alternative specifications agree about the timing of peaks and troughs, they disagree about the sizes of the output gap. We find that forecasts of inflation generated by these models improves 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 forecast inflation, and it is quite reliable in real time. Our estimate suggests that after the crisis the output gap is as large as -6% and our model correctly predicts falling inflation since 2012.

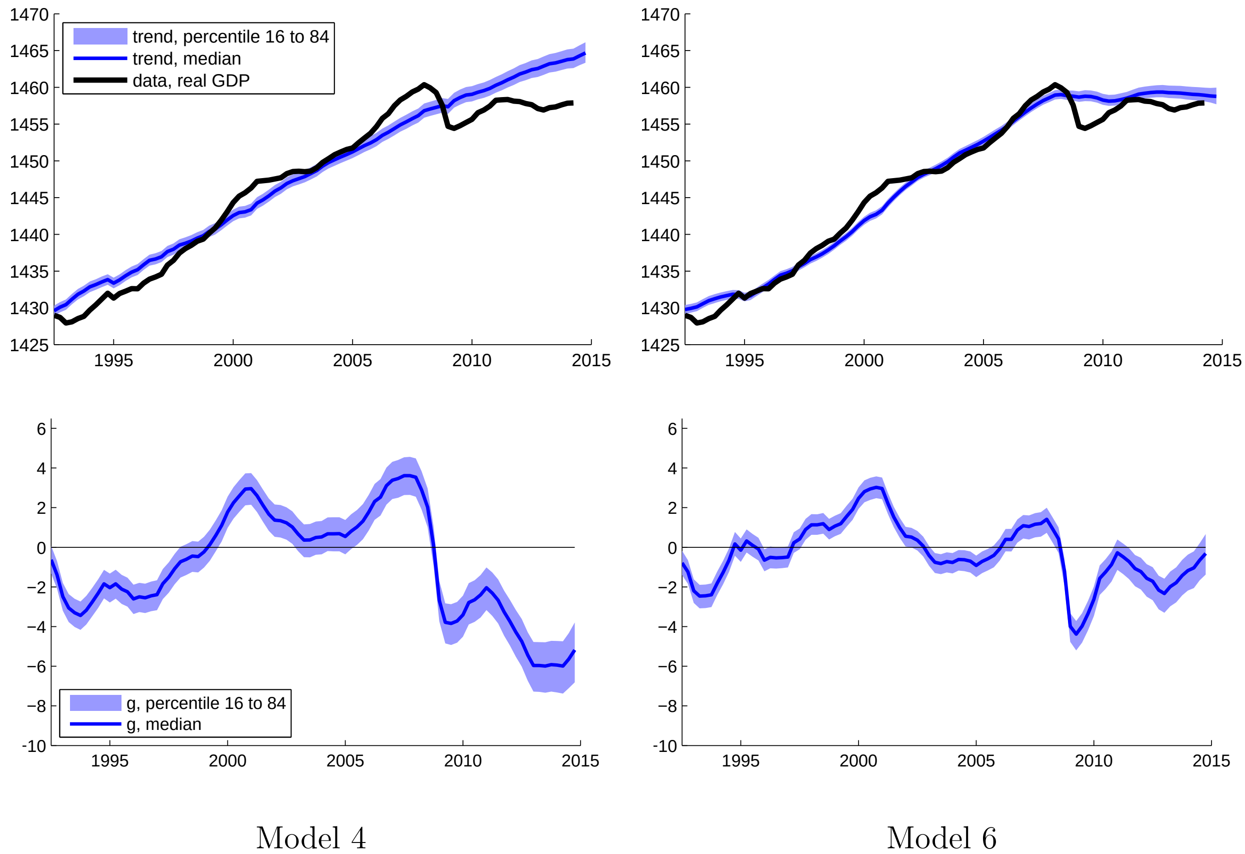



Figure 4: Trend output and output gap: random walk trend vs second order random walk trend

In principle, we could incorporate mixed frequency data to make our output gap even more timely, but we leave this extension for future research.

# 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 food | $p$ | log-diff ($\pi$) | $t-1$ | 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 | $y^5$ | raw | $t-1$ | Euro area RTD | 85Q1-90Q2 |
| Consumer confidence | $y^6$ | raw | $t-1$ | Euro area RTD | None |
| Capacity utilization | $y^7$ | raw | $t$ | Euro area RTD | None |
| Inflation expectations | $\pi^e$ | raw | $t$ 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 back in time using the growth rates from the Area Wide Model (AWM) database (Fagan et al. (2001)). Specifically, we take the level of the variable in the earliest available date in the RTD and we assume that before that date the variable evolved according to the growth rates of its AWM database counterpart. 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

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 prior mean of zero is a neutral benchmark. The variance is analogous to the variance of the Minnesota prior of Litterman (1986): the ratio $\tilde{\sigma}^2_{y^n_t}/\tilde{\sigma}^2_{y^1}$ accounts for the different volatilities of the left-hand-side variable $y^n_t$ and the right-hand-side variable $g_t$ (which is a component of real GDP). Notice that this 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 $\mathcal{N}(0, \tilde{\sigma}^2_{\pi}/\tilde{\sigma}^2_{y_1})$, analogously to the coefficients $B(L)$. The coefficients of $g_t$ and $g_{t+1}$ are fixed at 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 $\mathcal{N}(0, 0.1^2)$. Both inflation excluding energy and food $\pi_t$ and 5-year inflation expectations $\pi^e_t$ are measured in percentage points and we consider it likely a priori that they might differ by about 0.1 percentage point on average.

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

Trend inflation $z_t$ follows an AR(1) process and we center the prior at the values that imply the mean of 2% (consistent with the ECB definition of price stability) and moderate persistence, and we specify a rather large variance. In particular, 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 (e.g. 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 mean $d^z$ is $\mathcal{N}(0.4, 0.5)$. The value 0.4 in conjunction with the autoregressive coefficient of 0.8 implies the steady state of 2%.

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, $\tau$ 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.[7] The prior about $\tau$ 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 $\phi_1, \phi_2$. 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 $\tau, a$ implies certain dynamic properties of the output gap, formally summarized by the density

[7] 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.

$p(g)$. 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 interpretable.[8] Finding the prior for $\phi$, $p(\phi)$, means approximating the solution of the integral equation

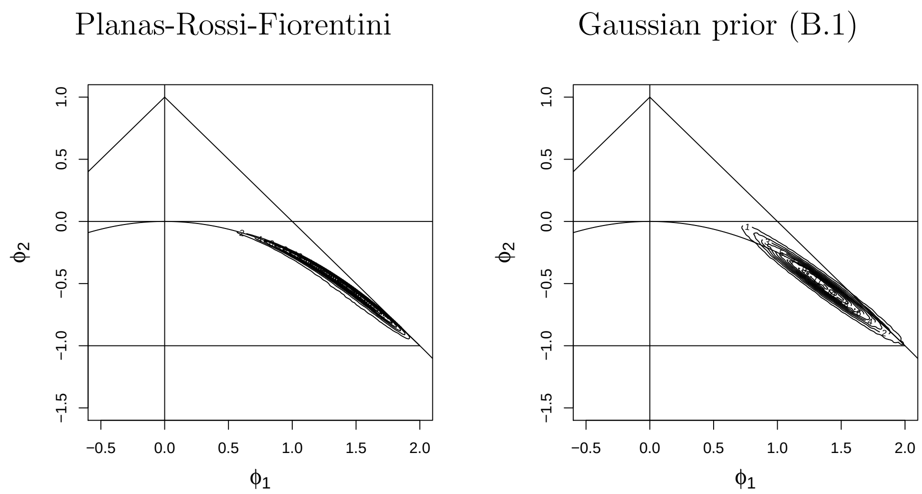
$$p(g) = \int p(g|\phi)p(\phi)d\phi \tag{B.3}$$

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

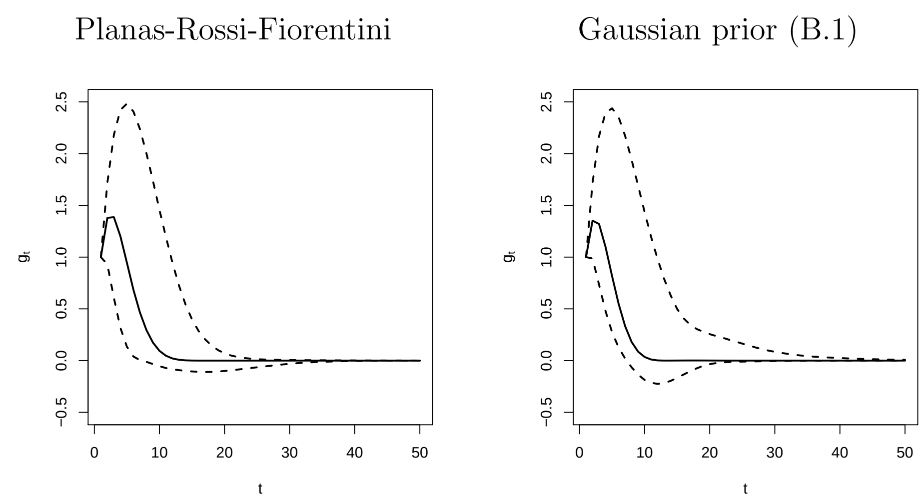
Figure B.1 illustrates the quality of the approximation. Panel A compares the densities of the coefficients $\phi_1$ and $\phi_2$ implied by (B.2) with the Planas-Rossi-Fiorentini prior (left plot) and Gaussian prior (B.1) (right plot). The Gaussian prior has 0.24 probability mass above the parabola $\phi_1^2+4\phi_2 = 0$, 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

[8]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 $69 \pm 46$ periods corresponds to $\rho \in (0.97, 0.994)$, i.e. a very different range for $\rho$. 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.




A. Joint densities of $\phi_1, \phi_2$. 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 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 half of the variance of $\Delta v_t$, and the variance of $\Delta v_t$ equals the training sample variance $\tilde{\sigma}_v^2$. We always refer to the variance of $\Delta v_t$ 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.

For all variables $y^n$, $n > 1$ (i.e., other than real GDP), the variances of the shocks in the trend equation $u_t^{w,n}$ and in the observation equation for $y_t^n$, $u_t^{y,n}$ have means respectively $\tilde{\sigma}_{y^n}^2/2$ and $\tilde{\sigma}_{y^n}^2/4$. To see that these means are consistent with our rule of thumb that half of the variance of $\Delta y^n$ is explained by the trend and half by the transitory shocks, note that at the prior mean $y_t^n = w_t^n + u_t^{y,n} = d^n + w_{t-1}^n + u_t^{w,n} + u_t^{y,n}$. Then $\Delta y_t^n = d^n + u_t^{w,n} + u_t^{y,n} - u_{t-1}^{y,n}$ and $var(\Delta y_t^n) = var(u_t^{y,n}) + 2var(u_t^{y,n})$. We follow the same rule of thumb in the remaining two observation equations: the prior mean of the variance of $u^\pi$ is $\tilde{\sigma}_\pi^2/4$ and the prior mean of the variance of $u_t^e$ is $\tilde{\sigma}_{\pi^e}^2/4$.

The prior mean of the variance of $u_t^g$ is $0.2\tilde{\sigma}_{y_1}^2$. This mean is consistent with the prior that, conditional on the prior means of $\phi_1$ and $\phi_2$, $g_t$ accounts for half of the variance of $\Delta y_t^1$. To see this, note first that $var(\Delta y_t^1) = var(u_t^{w,n}) + var(\Delta g_t)$ and $var(\Delta g_t) = \chi var(u_t^g)$ where $\chi$ is a function of $\phi_1$ and $\phi_2$. 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 $var(\Delta g_t) = \chi var(u_t^g) = 0.5\tilde{\sigma}_{y^1}^2$ we need to set $var(u_t^g) = 0.5/\chi\tilde{\sigma}_{y^1}^2$ and $0.5/\chi$ evaluates to about 0.2 when $\phi_1 = 1.352$ and $\phi_2 = -0.508$.

The prior mean of the variance of the shocks to trend inflation is $\tilde{\sigma}_{\pi^e}^2/2$.

**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 $\pi_0^e$. 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.

# 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.

Christiano, L. J., Trabandt, M., and Walentin, K. (2010). DSGE Models for Monetary Policy Analysis. In Friedman, B. M. and Woodford, M., editors, *Handbook of Monetary Economics*, volume 3 of *Handbook of Monetary Economics*, chapter 7, pages 285–367. Elsevier.

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.

Diron, M. and Mojon, B. (2008). Are inflation targets good inflation forecasts? *Economic Perspectives*, (Q II):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.

Galí, J., Smets, F., and Wouters, R. (2012). Unemployment in an Estimated New Keynesian Model. *NBER Macroeconomics Annual*, 26(1):329 – 360.

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. (1997). The time-varying nairu and its implications for economic policy. *Journal of Economic Perspectives*, 11(1):11–32.

Gordon, R. J. (2014). The turtle's progress: Secular stagnation meets the headwinds. In Coen, T. and Richard, B., editors, *Secular Stagnation: Facts, Causes and Cures*.

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. (2014). A note on implementing the Durbin and Koopman simulation smoother. MPRA Paper 59466, University Library of Munich, Germany.

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

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2013). Is There a Trade-Off between Inflation and Output Stabilization? *American Economic Journal: Macroeconomics*, 5(2):1–31.

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):25–38.

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.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 186119571. *Economica*, 25(100):283–299.

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.

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.