

How big is the output gap in the euro area?

Inflation will tell!*

Marek Jarociński[†] Michele Lenza[‡]

Preliminary, this version: December 19, 2015

Abstract

We estimate a small Bayesian dynamic factor model of the euro area, including a set of real activity variables and core inflation. Our measure of the output gap is the common factor underlying the cyclical fluctuations in the variables, normalized to coincide with the deviation of output from its trend. Different reasonable specifications of our empirical model yield very different estimates of the output gap. We discriminate among the models by evaluating the accuracy of their real-time inflation forecasts. The best model forecasts inflation well and implies that after the 2011 sovereign debt crisis the output gap in the euro area has been much larger than the official estimates imply. Models that feature a secular-stagnation-like slow-down in trend growth, and hence a small output gap after 2011, are rejected because they do not adequately capture the inflation developments.

*For comments and suggestions, we thank Carlo Altavilla, Marta Bańbura, Domenico Giannone, Siem Jan Koopman, Giorgio Primiceri, Lucrezia Reichlin 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.

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

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

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

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

1 Introduction

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

The approach developed 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 persistent trends while their fluctuations at business cycle frequencies, around their 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). We postulate that a policy-relevant measure of output gap is the one sending the most accurate real-time signal about future inflation. Therefore, we rank the different variants of our dynamic factor model according to the precision of out-of-sample, real-time forecasts of core inflation they produce. Our real-time exercise uses the 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 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 they correctly capture the fall in inflation. The output gap in the crisis is large and by 2014 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 slow-down of the trend growth (see e.g. Summers (2013), and Gordon (2014)). This hypothesis is widely debated and many other studies rather conclude that this recent weak growth reflects cyclical, although persistent, sources of fluctuation (see Blanchard (2015) for a recent survey of the arguments of the two sides). Our results highlight that reconciling a slow-down 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) 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. 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 and Modugno (2015) our model accounts for the presence 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.

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. In fact, the macroeconomic 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, collected 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 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 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

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

(which comes from 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.²

3 Econometric model

We use an unobserved components model to estimate output gap and to forecast inflation. The model relates the observed variables – real activity indicators y_t^1, \dots, y_t^N , inflation π_t , and inflation expectations π_t^e – to unobservable states. 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, \dots, w_t^N , and trend inflation z_t .

The observation equations of the model are

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

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

$$\pi_t^e = c + z_t + \varepsilon_t^e, \quad (1c)$$

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.

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

The state equations in the baseline version of the model are

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

$$w_t^n = d^n + w_{t-1}^n + \eta_t^n \quad \text{for } n = 1, \dots, N, \quad (2b)$$

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

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 y_t . 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 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.

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 .⁴

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.

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

⁴Often, in similar equations, also variables reflecting cost-push shocks are included, such as the oil price and the exchange rate. However, our measure of inflation is based on HICP excluding energy and unprocessed food and we found that including these cost-push variables did not improve our inflation forecasts.

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

$$\text{i.i.d. trend,} \quad w_t^n = d^n + \eta_t^n, \quad (3)$$

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

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

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

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

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

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 VARs. Here we only sketch the most relevant aspects, while the full details are reported in appendix B at the end.

⁵We 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.

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 with the Gibbs sampler.

3.2 Estimation

In the Gibbs sampler we draw the parameters ($b^n(L)$ for $n = 1, \dots, N$, $a(L)$, c , ϕ_1 , ϕ_2 , d , f , and all the shock variances) conditionally on the unobserved states (g_t , w_t^n for $n = 1, \dots, N$, z_t , all for $t = 1, \dots, T$), 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.

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.

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 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 a priori stationary

real activity variables (unemployment rate, consumer confidence and capacity utilization) are modeled as i.i.d. processes with a constant mean.⁶ In Models 6 and 7, the trends of the 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 most recent sample, 1992Q3 to 2014Q3. 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 year 2014 the estimates of the output gap range from above -2 to below -6 percent of GDP.

4.3 Forecasting results

In this subsection, we discriminate among the different measures of output gap by studying the real-time out-of-sample forecasting performance of our models. The design of the empirical validation exercise is as follows. We re-estimate each of the seven models over 55 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 $\pi_{t,t+h}$ defined as

$$\pi_{t,t+h} = \frac{400}{h} (p_{t+h} - p_t),$$

⁶The prior mean and variance of these constants are equal to the mean and variance of the corresponding variable in the training sample.

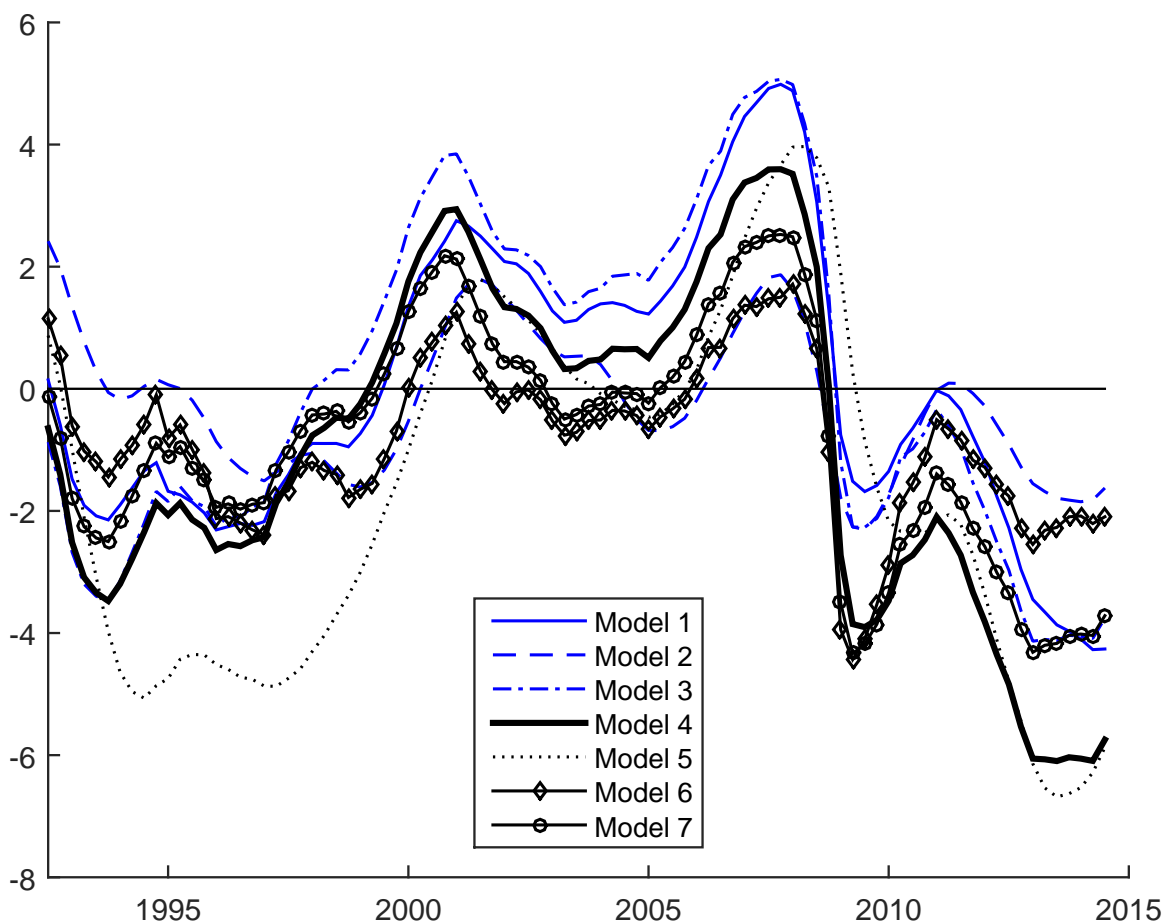


Figure 1 – Point estimates (posterior medians) of the output gap from Models 1 to 7

where p_t the log-level of consumer prices. We compute the target inflation rate using the latest available data, i.e. those available as of 2015Q3.

The first estimation sample (data available on 2001Q1) spans the period 1992Q3 - 2001Q1 and the last (data available on 2014Q4) 1992Q3 - 2014Q3. 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 $\pi_{t-1,t-1+h}|\mathcal{I}(t)$. 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, $\pi_{t-1,t}|\mathcal{I}(t)$, and one year ahead

forecasts, $\pi_{t-1,t+3}|\mathcal{I}(t)$. Our point forecasts are the medians of the posterior predictive densities. 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 .⁷ A number smaller than one indicates that the model outperforms the simple benchmark.

Table 2 – Inflation forecasts: real time performance.

	A. MSE of point forecasts (median)*		B. Log predictive density scores*	
	nowcast	1 year	nowcast	1 year
Model 1	0.60	0.73	0.32	0.52
Model 2	0.61	0.76	0.33	0.45
Model 3	0.58	0.66	0.33	0.48
Model 4	0.51	0.42	0.40	0.54
Model 5	0.58	0.44	0.34	0.35
Model 6	0.73	0.81	0.21	0.09
Model 7	0.69	0.65	0.20	0.19

* Normalized by the respective statistic from a naive benchmark model. Evaluation samples: 2001Q1-2014Q3 (nowcasts) 2002Q1-2015Q3 (1-year ahead forecasts). MSE: number smaller than one indicates that the model outperforms the naive benchmark. Log scores: number greater than zero indicates that the model outperforms the naive benchmark.

Panel A of Table 2 shows that the Phillips curve inflation forecasts (inflation forecasts using information on real activity) can outperform simple benchmarks in the euro area, 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 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,

⁷The random walk with drift for p_t is a standard simple benchmark and produces forecasts that are 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).

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. The final lesson we draw from the evaluation of the point forecasts is that allowing for the more flexible trend representations embedded in Models 6 and 7 does not pay in terms of forecasting accuracy, as these models are dominated by Models 4 and 5. 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.

Panel B of Table 2 corroborates these findings and complements the picture by reporting the log scores of predictive densities. In particular, panel B reports the average difference (over the sample) of log scores compared with the simple benchmark (positive values indicate the model outperforms the benchmark). Model 4 outperforms other specifications also when we evaluate the accuracy of the full predictive distribution. Interestingly, however, the relative rankings of the other models change, compared to the ranking in terms of point forecasts. Models 5, 6 and 7 are outperformed by Models 1 to 3 according to the log score, even though they produce better point forecasts. This is especially due to the different behavior of the predictive variances across models. In particular, the variances of the forecasts from Models 1, 2 and 3 tend to be larger. This helps mitigate the effects of forecast errors on the log score. Instead, Models 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. Model 4 provides the best compromise, given that it has both low variance and good precision of the point forecast.

Figure 2 presents one year ahead predictive densities of inflation along with the

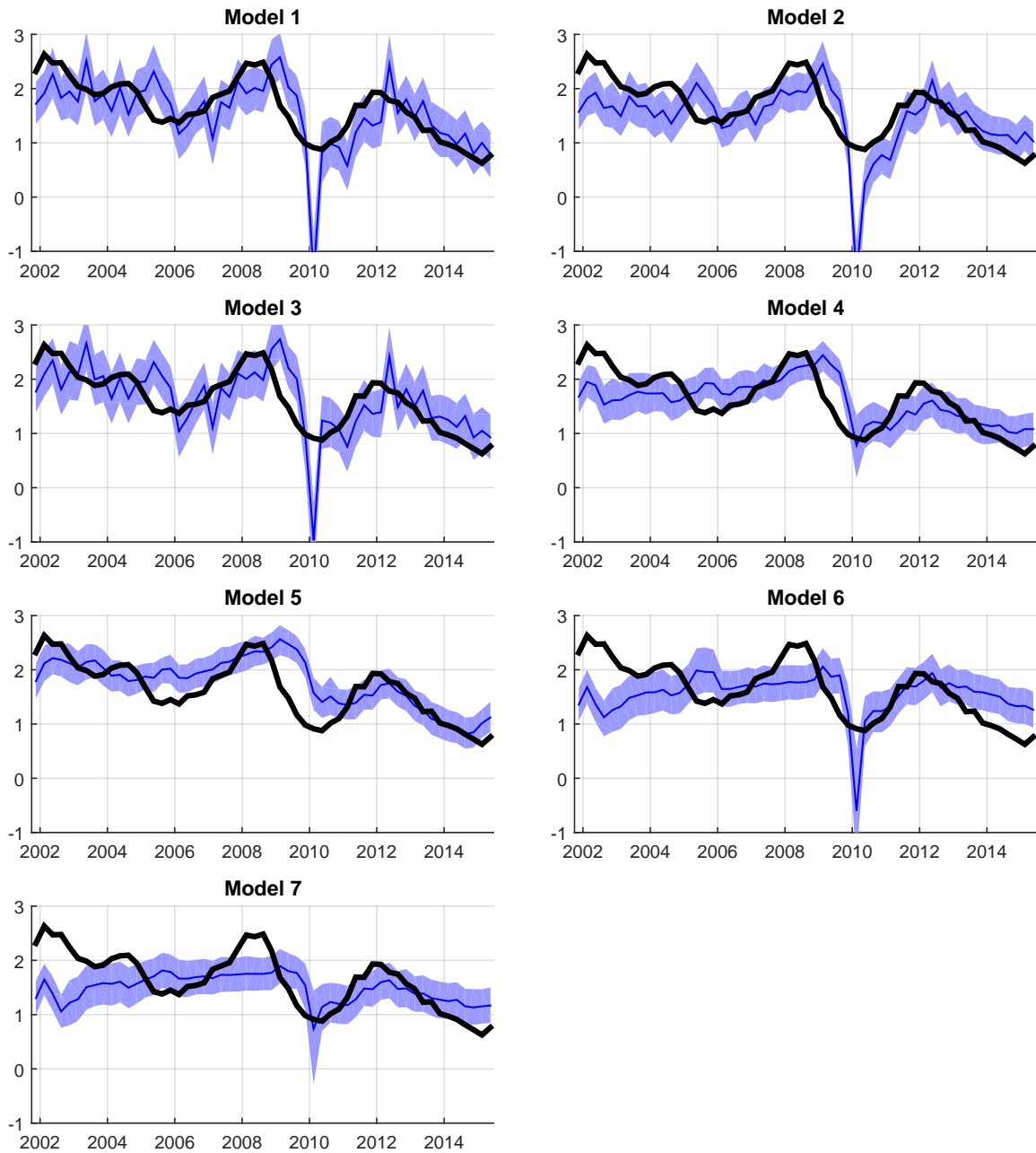


Figure 2 – One year ahead forecasts of inflation (blue line: median, blue shaded area: percentile 16 to 84) and actual inflation (black line)

actual inflation. The solid line is the observed annual inflation, final vintage, i.e. at time t it represents $100(p_t - p_{t-4})|_{\mathcal{I}(2015Q3)}$. 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 $100(p_t - p_{t-4})|\mathcal{I}(t-3)$. Two main lessons follow from this figure. First, it shows that one year ahead forecasts become much less volatile when (i) we use multiple indicators of real activity and (ii) 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 7. Second, the forecasts from Models 4 and 5 are rather similar and track inflation better than the forecasts from Models 6 and 7, which are also similar to each other.

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 55 real time posteriors. More in details, at each date we have up to 55 sets of posterior quantiles of the output gap, obtained with our 55 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 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

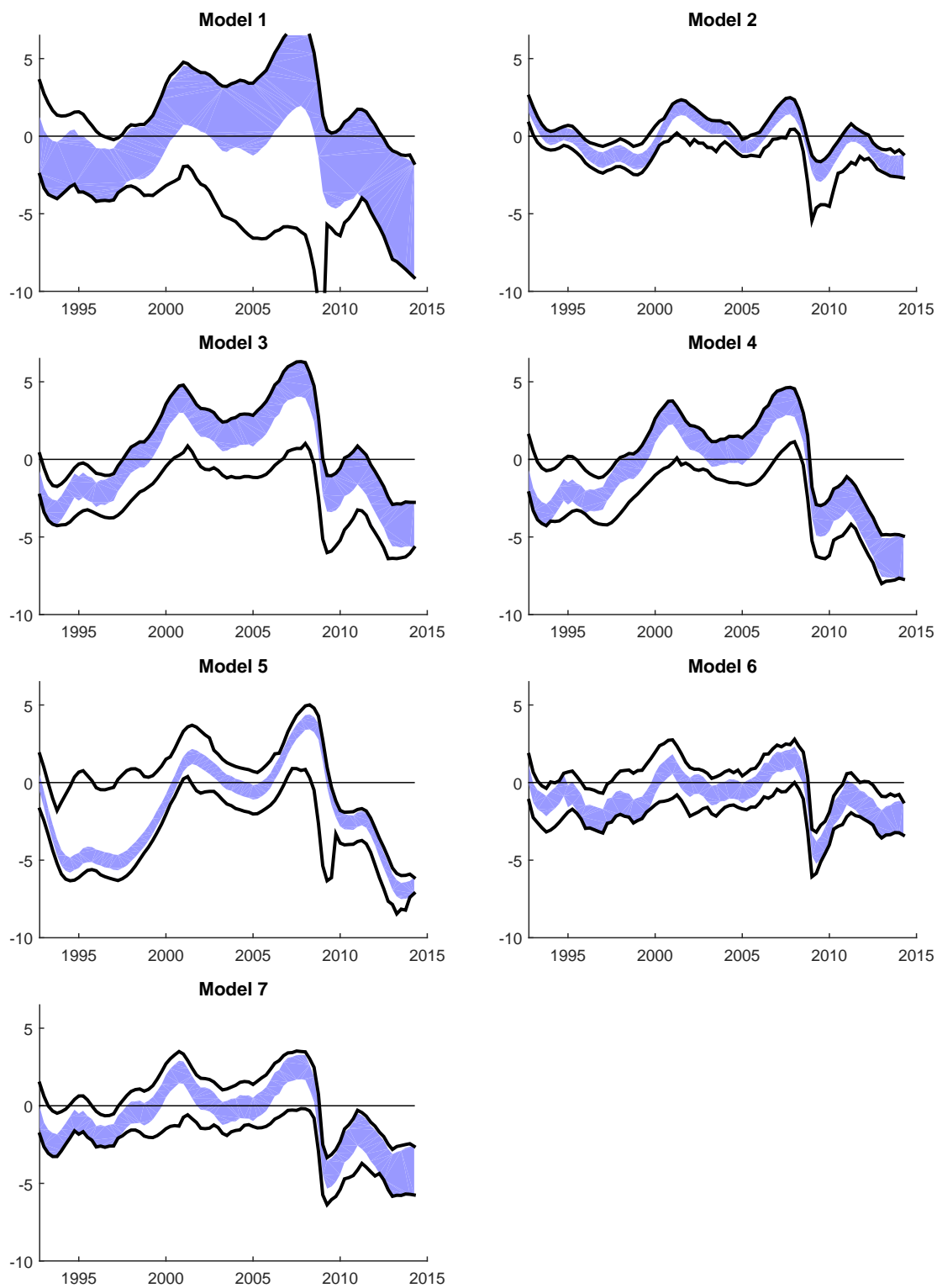


Figure 3 – 16th and 84th percentiles of the output gap: envelope of all the real time samples (black line) and the last sample (blue shaded area)

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.⁸ Hence, 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.

4.5 The output gap in the current crisis

According to our best performing models, in the last part of the recent double-dip recession we observe the largest (negative) output gap in the history of the euro area. This finding has important policy implications. Taken at face value, the finding of a large output gap suggests 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 illustrates these alternative views. 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. Model 6 is consistent with the view that trend output growth has stalled. Model 7 is in-between. The official output gap estimates by the IMF and the European Commission, also shown in this Figure, are between those obtained by Model 6 and 7.

Figure 4 makes it clear that the properties of the trend output are crucial. In

⁸Mertens (2014) shows that adding stochastic volatility to a small model like Model 1 strongly improves the real-time reliability of output gap estimates. We leave it for future research to study whether stochastic volatility can further improve the real-time reliability in larger models like Model 4.

Model 7, the growth rate in trend output changes little and in Model 4 not at all. Then, when economic growth stalls, as it has since the beginning of the recession in the euro area, the output gap opens ever wider. In particular, in Models 4 and 7 the output gap in the second dip is larger than it was in the first dip. By contrast, in Model 6 growth rate of trend output falls sharply, and trend output tracks more closely than in other model the developments in actual output. As a consequence, in Model 6 and in 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 very hard to reconcile the persistently low inflation with small output gaps. We find that Model 4 (as well as Models 5 and 7) 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 institutions, whose output gap estimates resemble a lot those of model 6, were excessively high after 2012.⁹

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 coincide about the timing of peaks and troughs, they disagree about the size of the output gap. We find that the 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

⁹Our 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.

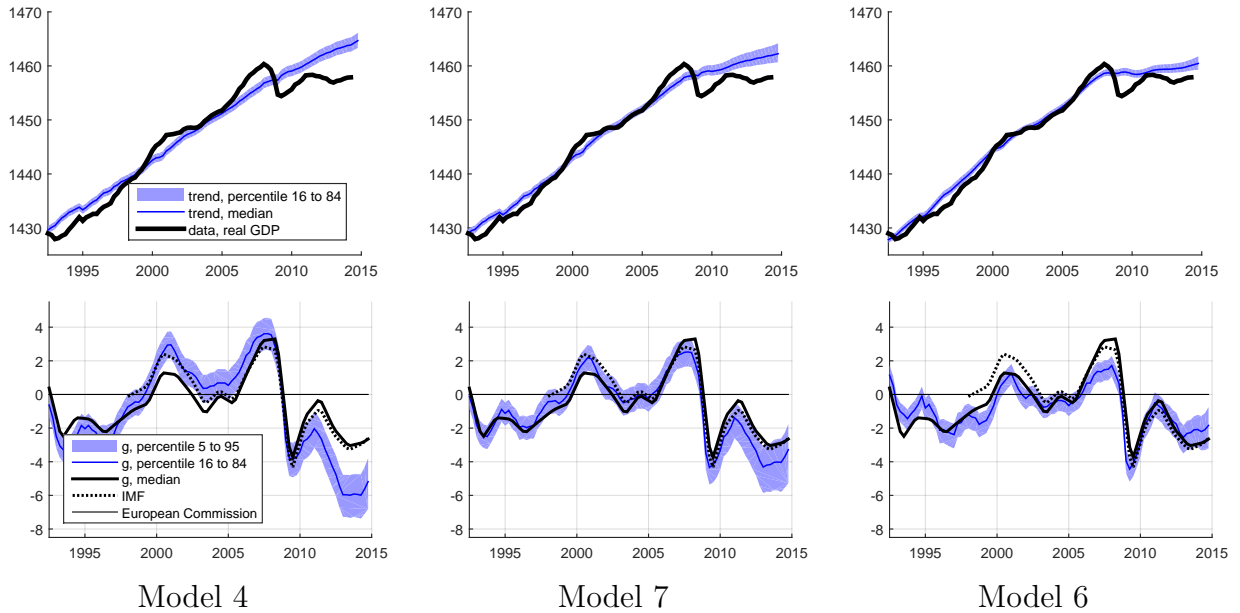


Figure 4 – Trend output and output gap according to Models 4, 6, 7, IMF and European Commission

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 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.

We leave for future research several possible extensions of our work. In particular, incorporating mixed frequency data (for example, monthly variables like industrial production and business surveys) might make our output gap estimates even more timely and precise in real-time. Moreover, adding stochastic volatility could further improve the real-time reliability of output gap estimates and the density forecasts of inflation, as e.g. in Mertens (2014). It might be worth also to explore whether combining the forecasts from our suite of models yields substantial gains.

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	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	π^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 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 models.

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 $v \in \{y_t^1, \dots, y_t^N, \pi_t, \pi_t^e\}$ we compute the mean $\tilde{\delta}_v = \frac{1}{T^{tr}-1} \sum_{t=2}^{T^{tr}} \Delta v_t$ and variance $\tilde{\sigma}_v^2 = \frac{1}{T^{tr}-1} \sum_{t=2}^{T^{tr}} (\Delta v_t - \tilde{\delta}_v)^2$.

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 $\mathcal{N}(0, \tilde{\sigma}_{y_t^n}^2 / \tilde{\sigma}_{y_t^1}^2)$. 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 $\tilde{\sigma}_{y_t^n}^2 / \tilde{\sigma}_{y_t^1}^2$ 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 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 $\mathcal{N}(0, \tilde{\sigma}_\pi^2 / \tilde{\sigma}_{y_t^1}^2)$, analogously to the coefficients $b^n(L)$. 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 unprocessed food π_t and long-term inflation expectations π_t^e 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_t^n is a random walk with drift, $\Delta w_t^n = d^n + \eta_t^n$. The drift d^n is $\mathcal{N}(\tilde{\delta}_{y_t^n}, \tilde{\sigma}_{y_t^n}^2)$ 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 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 for the constant term 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

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

To arrive at this prior we start with the auxiliary model

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

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.¹⁰ 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

¹⁰The prior is $(\tau - 2)/(141 - 2) \sim \text{Beta}(2.96, 10.70)$ and $a \sim \text{Beta}(5.82, 2.45)$, see Planas et al. (2008), p.23.

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 interpretable.¹¹ 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 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, (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 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

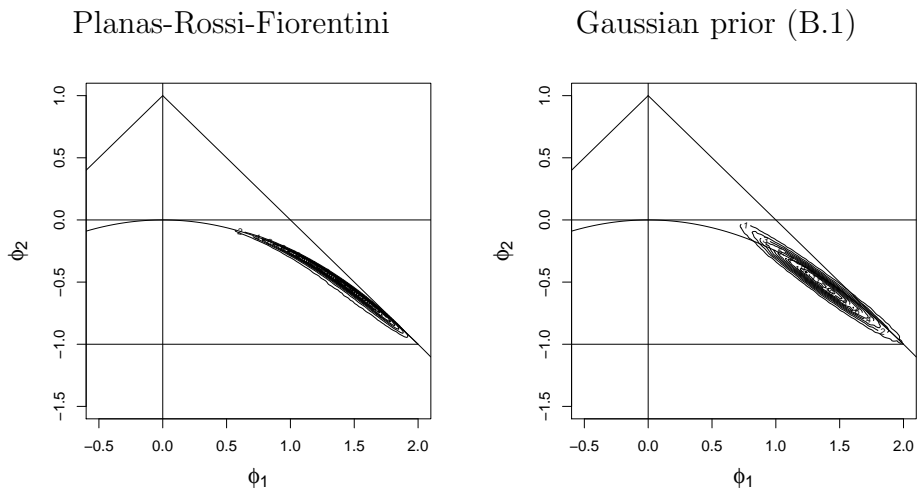
¹¹To see how important it is to think in terms of the behavior of the modeled variable and not in terms of model parameters, think of the following illustrative example. Consider a process x_t and a model $x_t = \rho x_{t-1} + \varepsilon_t$. Suppose one's prior on the half-life of x_t is centered at 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.

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.

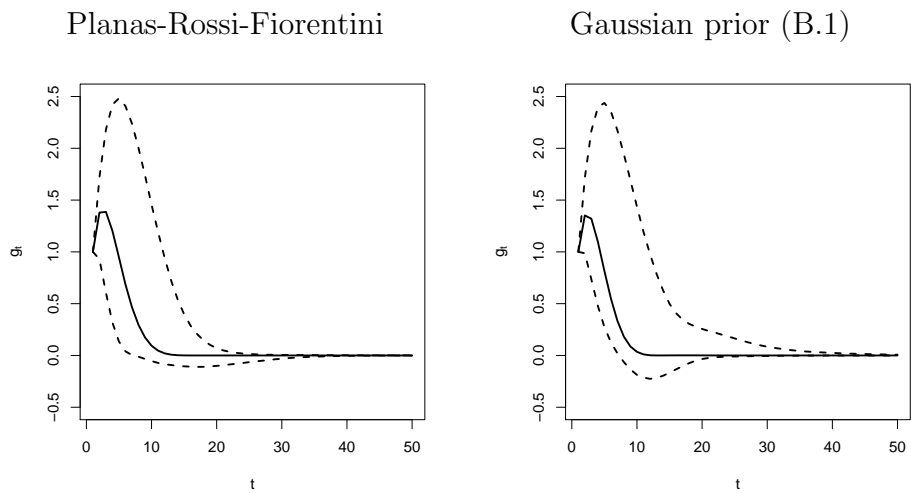
Shock variances. When setting the priors about the variances of the shocks we use the rule of thumb that for each observable series v_t , when all the coefficients are at their prior means, the trend and non-trend components account a priori for a half of the variance of Δv_t each, and the variance of 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.

For all variables y^n , $n > 1$ (i.e., other than real GDP), the variances of the shocks in the trend equation η_t^n and in the observation equation for y_t^n , ε_t^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 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 $var(\Delta y_t^n) = var(\eta_t^n) + 2var(\varepsilon_t^n)$. 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 $var(\Delta y_t^1) = var(\eta_t^g) + var(\Delta g_t)$ and $var(\Delta g_t) = \chi var(\eta_t^g)$ where χ is a function of ϕ_1 and ϕ_2 . It is straightforward, though tedious, to show that $\chi = 2(1 - \phi_1 - \phi_2) / ((1 + \phi_2)((1 - \phi_2)^2 - \phi_1^2) + 1)$. See e.g. Hamilton (1994) pp.57-58 for similar derivations. Hence, if we want $var(\Delta g_t) = \chi var(\eta_t^g) = 0.5\tilde{\sigma}_{y^1}^2$ we need to set $var(\eta_t^g) = 0.5/\chi \tilde{\sigma}_{y^1}^2$ and $0.5/\chi$ evaluates to about



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

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 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 model comparison exercise which, for brevity, we omitted from the main body of the text and report it here. We try a grid of 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. Table B.1 reports the mean squared forecast errors. The differences in the MSEs across models are very small. It turns out that in Model 7 there is a trade-off between avoiding a large error at the beginning the crisis and matching the post-2012 disinflation. Different models strike a different balance between these two indicators so that the overall MSEs differ little. More flexible models make small mistakes in the beginning of the crisis and then miss the post-2012 disinflation, and vice versa for more rigid models. Another observation is that although in the final sample Model 7d produces a similar output gap as Model 4 (see Figure B.2), but in real time it is often quite different. This is manifested in a large difference between the MSEs of Model 7d in Table B.1 and those of Model 4 in Table ???. Version c tends to produce the lowest MSEs at longer horizons, and it is also nontrivially different

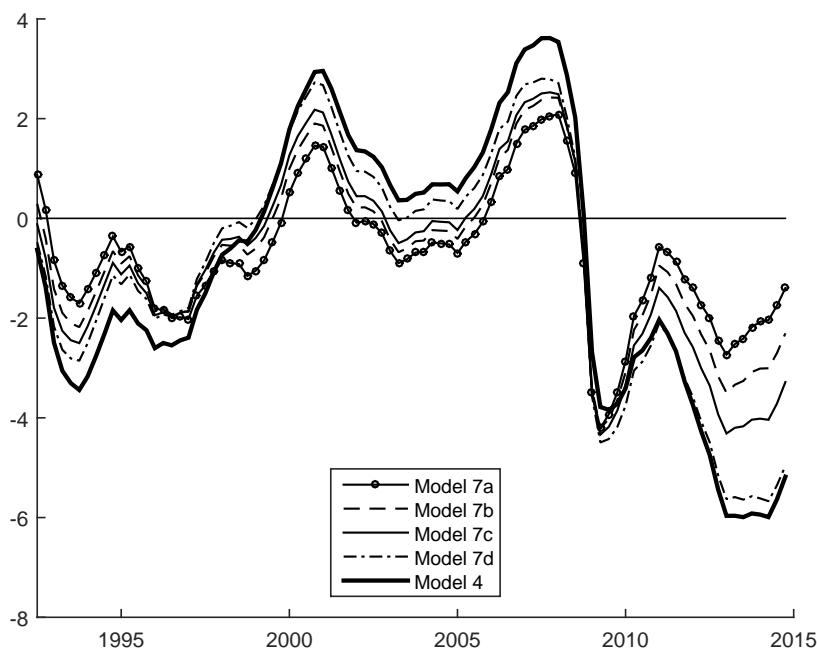


Figure B.2 – Posterior medians of the output gap from different versions of Model 7

from Model 4, and therefore we pick it to be reported in the main body of the paper as simply Model 7.

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

Table B.1 – MSE relative to the simple benchmark: different variants of Models 6 and 7.

	nowcast	1 year	2 years		nowcast	1 year	2 years
Model 6a	0.681	0.791	1.014	Model 7a	0.658	0.653	0.695
Model 6b	0.726	0.839	1.020	Model 7b	0.681	0.653	0.627
Model 6c	0.731	0.815	0.970	Model 7c	0.686	0.647	0.606
Model 6d	0.770	0.785	0.803	Model 7d	0.697	0.647	0.591

Note: Ratio of mean squared forecast error of variants of Models 6 and 7 relative to the random walk with drift for p_t . Current quarter ($h=0$) and four quarters ahead forecasts ($h=4$). Numbers smaller than one indicate that the model outperforms the random walk.

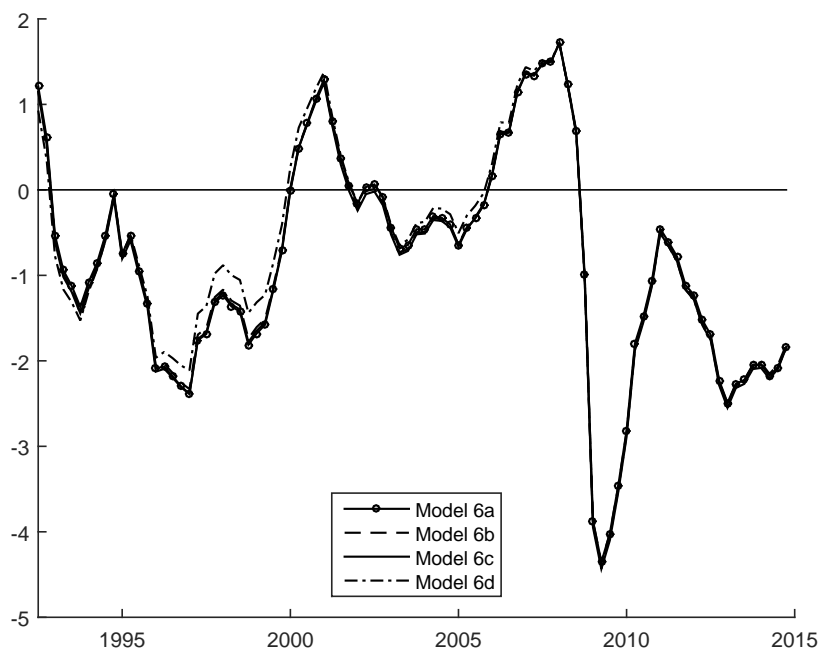


Figure B.3 – Posterior medians of the output gap from different versions of Model 6

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 the variables of interest 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. The variables of interest reported in Table C.1 include all the parameters of Model 4 except for the coefficients $b^n(L)$ (which are discussed below), as well as the output gap g_T and one-year-ahead inflation forecast $\pi_{T,T+4}$ at the end of the last vintage ($T=2014Q4$). 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.

The values of the Geweke's Z statistics for these coefficients (unreported for brevity) range from 0.21 to 1.58, so we never reject convergence at 5% level. The RNE's range from 0.18 to 0.77.

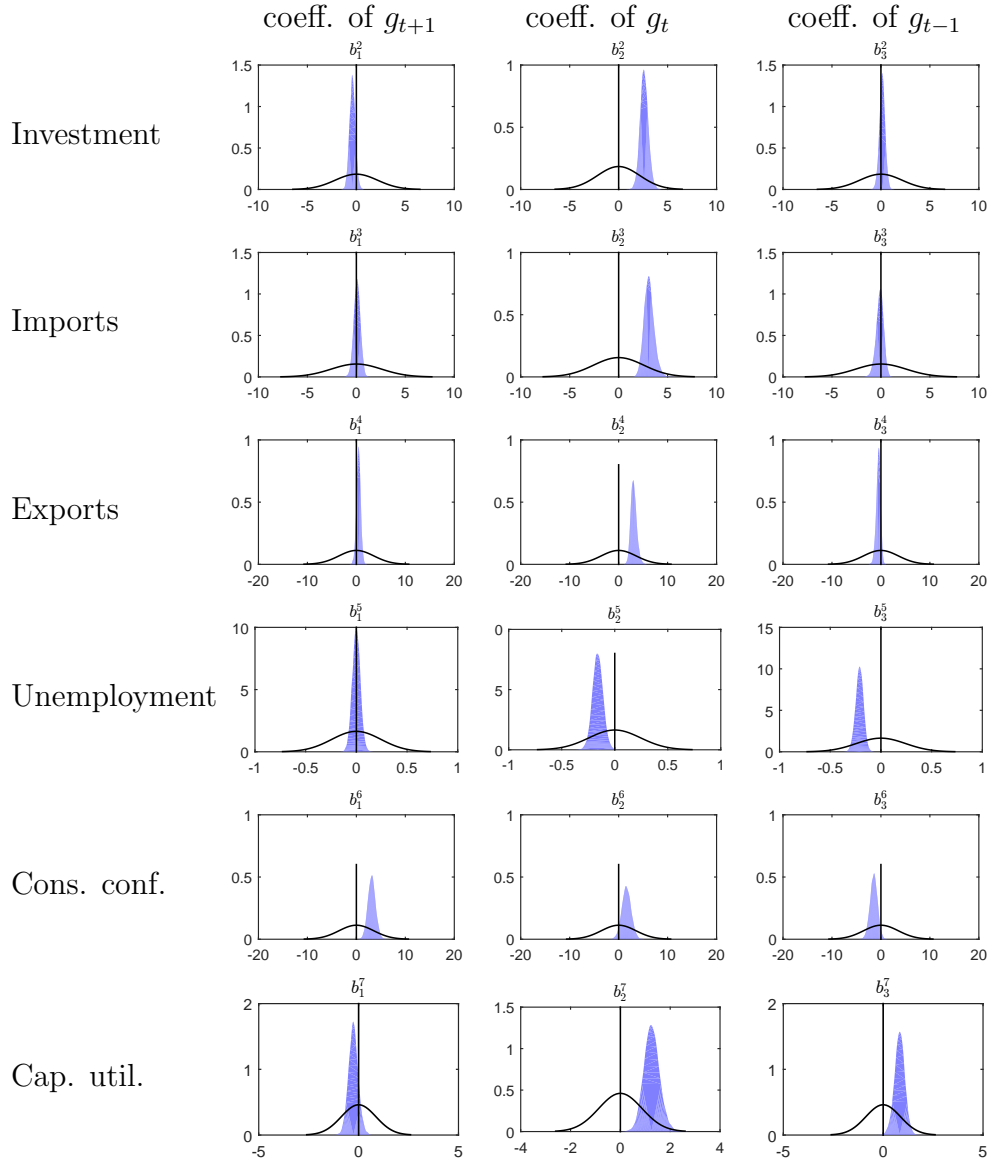


Figure C.1 – Prior and posterior densities of the coefficients of g in the observation equations of real activity variables.

Table C.1 – Posteriors and convergence diagnostics

var	Mean	Sd.	p1	p50	RNE	Z	$p(z > Z)$
logsig2y2	-7.314	0.409	0.07	-0.02	1.09	1.67	0.09
logsig2y3	-2.537	0.468	0.15	0.02	0.67	0.56	0.58
logsig2y4	-1.543	0.500	0.18	-0.01	0.97	0.82	0.41
logsig2y5	-4.175	0.557	0.23	0.02	0.78	0.12	0.90
logsig2y6	-1.974	0.380	0.04	0.01	1.15	0.13	0.90
logsig2y7	-2.347	0.420	0.08	0.00	0.93	1.33	0.18
c1	0.188	0.097	0.96	0.14	0.13	0.37	0.71
logsig2x1	-6.500	0.544	0.22	0.01	0.87	1.14	0.25
phi1	1.726	0.071	0.06	0.02	0.61	1.35	0.18
phi2	-0.756	0.070	0.04	0.02	0.75	1.19	0.23
d1	0.396	0.029	0.06	0.00	0.45	0.18	0.86
d2	0.000	0.000	0.00	0.00	1.31	0.31	0.76
d3	0.359	0.083	0.06	0.01	0.54	0.14	0.89
d4	0.000	0.002	0.00	0.00	1.01	1.02	0.31
d5	0.000	0.001	0.01	0.01	0.93	0.41	0.68
d6	1.381	0.116	0.04	0.00	0.63	0.55	0.58
d7	1.264	0.094	0.05	0.00	0.53	0.61	0.54
d8	0.122	0.041	0.01	0.00	0.66	0.85	0.40
f1	0.928	0.021	0.01	-0.01	0.99	0.87	0.38
logSig1	-2.062	0.225	0.24	0.05	0.26	0.34	0.73
logSig2	-2.899	0.229	0.20	0.05	0.29	0.23	0.82
logSig3	-4.343	0.198	0.13	0.01	0.39	0.17	0.86
logSig4	-0.772	0.244	0.07	-0.01	0.52	1.28	0.20
logSig5	1.618	0.171	0.03	0.00	0.67	0.44	0.66
logSig6	-0.595	0.174	0.08	0.00	0.43	0.43	0.67
logSig7	-0.065	0.217	0.06	0.02	0.56	0.61	0.54
logSig8	-0.604	0.253	0.06	0.00	0.49	0.58	0.56
logSig9	-5.303	0.259	0.03	-0.02	0.97	0.94	0.35
az1	1.002	0.010	0.01	0.01	0.69	0.34	0.74
ag1	-0.006	0.010	0.00	0.00	1.04	0.91	0.36
ag2	-0.003	0.010	0.00	-0.02	1.04	0.41	0.68
ag3	0.120	0.028	0.09	0.04	0.29	0.02	0.99
logsig2pi	-1.352	0.159	0.03	-0.02	0.83	0.96	0.34
g_T	-5.808	1.368	0.40	0.04	0.21	0.59	0.55
$\pi_{T,T+4}$	1.083	0.312	0.01	0.00	0.69	0.96	0.34

Note: Statistics based on 10,000 draws from the Gibbs sampler. Mean: mean of the draws; Sd.: standard deviation of the draws; p_j : 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 $-Z$.

Appendix D 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 D.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 D.1 – Marginal likelihood of Models 4, 5, 6 and 7, last vintage.

Model number and trend process	log marginal likelihood
Model 4: RW	-860
Model 5: RW or i.i.d	-1030
Model 6: IRW	-905
Model 7: LLT	-845

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., Çakmaklı, 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.
- Chan, J. C., Clark, T., and Koop, G. (2015). A new model of inflation, trend inflation, and long-run inflation expectations. Technical report.
- 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, Antonello, D. G. M. L. 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*, (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.
- 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*.
- 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.
- 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.
- 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.
- Summers, L. (2013). Why stagnation might prove to be the new normal. Lawrence Summers’ blog, December 15, 2013.
- 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.