EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Marta Bańbura, Elena Bobeica

Does the Phillips curve help to forecast euro area inflation?

No 2471 / September 2020

Abstract

We find that it does, but choosing the right specification is not trivial. We unveil notable model instability, with breaks in the performance of most simple Phillips curves. Euro area inflation was particularly hard to forecast in the run-up to the EMU and after the sovereign debt crisis, when the trend and for the latter period, also the amount of slack, were harder to pin down. Yet, some specifications outperform a univariate benchmark most of the time and are thus a useful element in a forecaster's toolkit.

We base these conclusions on an extensive forecast evaluation over 1994 - 2018, an extraordinarily long period by euro area standards. We complement the analysis using real-time data over 2005-2018.

As lessons for practitioners, we find that: (i) the key type of time variation to consider is an inflation trend; (ii) a simple filter-based output gap works well overall as a measure of economic slack, but after the Great Recession it is outperformed by endogenously estimated slack or by estimates from international economic institutions; (iii) external variables do not bring forecast gains; (iv) newer generation Phillips curve models with several time-varying features are a promising avenue for forecasting, especially when density forecasts are of interest, and finally, (v) averaging over a wide range of modelling choices offers some hedge against breaks in forecast performance.

JEL Classification: C53, E31, E37

Keywords: Forecasting, Inflation, Phillips curve, Detrending, Time-varying parameters, Density forecast

Non-technical summary

The Phillips curve links inflation to economic activity. It is a key macroeconomic relationship, which has traditionally been used to understand inflation's past and future. There have been times when its validity came under scrutiny and this occurred especially in the aftermath of the Great Recession. Economists were puzzled first by the relative resilient inflation rates in countries with large output losses and afterwards by the prolonged low inflation in spite of the economic recovery.

This paper speaks to this debate and assesses the viability of the Phillips curve through its usefulness in forecasting inflation. We find that the Phillips curve does exist and is useful in forecasting euro area inflation, but choosing the appropriate specification is far from trivial. We base this result on an investigation of a large set of Phillips curve models, also considering 'newer generation models'. We conduct the forecast evaluation over 25 years (1994-2018), being the most comprehensive assessment of the Phillips curve forecasting performance for the euro area to the best of our knowledge. We complement the analysis with a real-time forecast evaluation (albeit on a shorter sample) that confirms our results.

More in detail, we take a simple stylised Phillips curve model and evaluate whether its ability to forecast euro area Harmonized Index of Consumer Price (HICP) inflation excluding energy can be improved by: (i) accounting for a time-varying inflation trend; (ii) changing how economic activity (or slack) is measured; (iii) including external drivers of inflation and finally (iv) adopting more recent econometric approaches that allow for general forms of time variation in the relationship and for deriving (within-model) measures of economic slack. We draw important lessons for the practitioners trying to forecast inflation.

Our findings are similar to some which were found in the US literature and bring new insights pertaining to the euro area. We find that it is indeed *hard to beat a univariate model* of inflation. Yet, some Phillips curve specifications can offer forecasts improvements most of the time, albeit limited. In other words, measures of economic activity do help to forecast inflation.

Unlike some studies for the US, we do not find a link between the times in which Phillips curves fail to beat a univariate benchmark model to certain states of the economy, namely expansion versus recession. Instead, euro area inflation was particularly hard to forecast in the run-up to the Economic and Monetary Union (EMU) and in the aftermath of the sovereign debt crisis. In both periods, it was particularly challenging for forecasters to capture the inflation trend and in the latter period, also the amount of slack in the economy.

One important result that we find is that adjusting for a time-varying trend in inflation is the key type of time variation to consider. A trend measure based on a weighted average of past inflation (with higher weight for more recent outcomes) appears to be a good choice for

estimating the inflation trend, as the gains with respect to the no detrending version prevail through most of the evaluation sample. Long-term survey measures of inflation expectations also tend to work well, but not over the latest low inflation period. The accelerationist version of the Phillips curve, i.e. a specification in which differences rather than levels of inflation are linked to economic slack, performs well in the pre-EMU inflation convergence period.

Our paper also looks into the appropriate measure of economic slack when trying to forecast inflation. We employ a wide set of product and labour market slack indicators, some of which have been constructed and back-casted for this investigation; we also explore a novel dataset of real-time estimates of the output and unemployment gaps published by international economic institutions (the OECD, the IMF, the European Commission and the Eurosystem/ECB). We find that the output gap obtained by simply filtering (log) real GDP performs relatively well compared to all alternatives. On the labour market it is harder to single out one indicator that is superior over the entire period. Especially in the aftermath of the Great Recession alternative indicators to the conventional filter-based unemployment gap bring some forecast gains, perhaps related to structural changes in the labour market. The changing configuration of the labour market in terms of increasing informal jobs and other non-standard work agreements renders the estimation of labour market slack more problematic. While in general making use of the slack estimates from the international economic institutions does not provide a 'silver bullet' when forecasting with Phillips curves, they do seem to bring forecast gains in this particular period, after the euro area double dip recession.

Further, we find that external drivers do not enhance the out-of-sample predictive power of the Phillips curve. We link this result with the difficulty in forecasting such variables and admit that external supply-side shocks can have a considerable explanatory power for domestic inflation in-sample.

New generation Phillips curve models (such as the one proposed by Chan et al., 2016, 2018) offer forecasting gains relative to a univariate benchmark and their main advantage appears to be related to the incorporation of time-varying trends and within-model estimation of economic slack.

The best performing models over the last few years are typically those with a low inflation trend. The analysis does not allow to draw conclusions on the potential drivers of a decline in the inflation trend. They could relate to the credibility of the monetary authority, demographics, financial factors, the rise of globalisation or of e-commerce. These are important questions for future research.

Overall, there is evidence of model instability, especially for simple models. Averaging forecasts, e.g. across functional forms of the inflation-slack relationship, estimation windows and lag selection criteria, offers some hedge against breaks in the forecast performance.

1 Introduction

It would be extraordinarily useful to discover a specification of the Phillips curve that fits the data reliably ... as Stock and Watson (2010) observe, the history of the Phillips curve 'is one of apparently stable relationships falling apart upon publication.' Ball and Mazumder (2011) is a poignant example. Nonetheless, because of the practical importance of the Phillips curve, researchers must continue to search for better specifications. (Ball and Mazumder, 2019a)

The Phillips curve has been long used to understand inflation's past and future and guide the conduct of monetary policy.[1] Lately, this relationship has come under intense scrutiny, with some economists doubting its usefulness altogether. A heated debate on weather the Phillips curve is *dead* or *alive* marked the post Great Recession period, with fundamental implications on how central banks are doing business (see the strategy review undertaken by the Federal Reserve System, where the validity of the Phillips curve has taken center stage, as discussed by Powell, 2020). The trigger of the discussion has been the puzzling flatness of the Phillips curve after the Great Recession. First, the inflation rate was quite resilient when many countries were facing large output losses; afterwards, in spite of the economic recovery, inflation appeared to be surprisingly low (see e.g. Blanchard, 2016; Hooper et al., 2019, and references therein for a more detailed discussion).

Different proposals were put forward in order to reconcile the path of inflation with that of the real activity. These proposals include: (i) looking at *an alternative measure of economic slack* (which is the proxy for the real marginal cost) such as the short-term unemployment rate (Gordon, 2013), the unemployment recession gap (Stock and Watson, 2010), a broad labour underutilisation measure (Bell and Blanchflower, 2018), or a model-determined ("endogenous") measure (Stella and Stock, 2015; Chan et al., 2016; Jarociński and Lenza, 2018); (ii) considering the *role of oil prices* (Coibion and Gorodnichenko, 2015; Hasenzagl et al., 2018; Coibion et al., 2019); (iii) allowing for the *Phillips curve slope to vary* (e.g. IMF, 2013; Blanchard et al., 2015; Constancio, 2015; Stella and Stock, 2015; Ciccarelli and Osbat, 2017); (iv) *switching from an "accceleretionist" to a "level" relationship* between inflation and slack (e.g. Blanchard, 2016; Ball and Mazumder, 2019a); (v) or properly accounting for the *changing trend* in inflation (Hasenzagl et al., 2019).[2]

In this paper we assess whether the Phillips curve is still alive through its usefulness for forecasting euro area inflation. We take the elements put forward in the aforementioned studies, many of which focus on understanding inflation in-sample, and investigate whether they are also useful when it comes to forecasting.

In particular, we are interested whether and when the Phillips curve can produce more accurate forecasts than those from popular univariate benchmarks proposed by Atkeson and Ohanian (2001) and Stock and Watson (2007). Given the aforementioned (and other) proposals, there is a rich space of possible Phillips curve specifications. We run a systematic comparison of a large number of specifications and investigate how various elements affect the forecast performance. We start by taking a simple workhorse model and evaluate whether it can be improved by: (i) accounting for a time-varying inflation

[1] Constancio (2015), Yellen (2015), Powell (2018) and Lane (2019) are among the policy makers extensively referencing the Phillips curve framework

[2] Some authors also propose alternative measures of inflation, (see e.g. Ball and Mazumder, 2019b; Stock and Watson, 2019). We do not pursue this idea here.

trend; (ii) changing how economic slack is measured; (iii) including external variables and finally (iv) adopting more recent econometric approaches ('new generation' Phillips curve models) with features such as time-varying parameters, stochastic volatility or model-determined measures of slack. We evaluate how the forecast performance relative to the benchmark models has evolved over time, which is a key element, as better performance of Phillips curves has been shown to be episodic in earlier studies (see e.g. Fisher et al., 2002; Stock and Watson, 2010). For practitioners, the stability of the forecast performance is key, it shows whether they can rely on models that worked well in the past to predict the future. Finally, we also assess whether the conclusions are robust to certain modelling choices such as the "functional form" of the relationship, the estimation window or the lag selection approach. The paper thus focuses on exploring the empirical relationship between inflation and activity for forecasting purposes; it does not go into identification issues in a Phillips curve, for instance (as discussed in McLeay and Tenreyro, 2019; Del Negro et al., 2020).

Admittedly, the empirical literature dealing with forecasting with Phillips curves is large (see Stock and Watson, 2009, for a comprehensive survey), but most of the lessons are derived for the US case. In addition, some of the recent proposals have not been systematically compared in terms of forecasting performance. Last but not least, extending the evaluation sample with the "missing inflation" episode has potential to offer new insights.

We focus on the euro area, for which evidence is much scarcer, albeit growing. The available studies often have a different focus, are based on a shorter evaluation sample, consider a narrower set of model specifications, and/or rely on in-sample analysis.[3] The euro area is an interesting case for various reasons. First, it has experienced two severe recessions in the recent past and stubbornly low inflation in the recovery phase (below the rates seen in the US). Second, it has particular structural features which differ from the US related to, e.g., rigidities of the labour market; it has also undergone massive structural reforms, which could have affected the inflation - output relationship. Last but not least, due to euro area's short history and related data limitations, an investigation of the forecasting performance over a long period of time has been challenging. To the best of our knowledge, this is the first paper to offer a comprehensive assessment of the Phillips curve forecasting performance for the euro area over such a long period, namely 25 years.

Our investigation provides important lessons for practitioners in terms of what works and what doesn't when specifying a Phillips curve model for forecasting. We believe this is useful as taking the simple idea underlying the Phillips curve to the data is not trivial and the exact specification differs with almost every published paper. Our conclusions can be summarised as follows.

First, similarly to the US case, instability seems to generally characterize the forecasting performance of the Phillips curve. While this is true, some Phillips curve specifications can offer improvements, even if modest, over a univariate model most of the time. In other words, measures of economic activity do help to forecast inflation. Unlike Stock and Watson (2010), we do not find a link between the times in which Phillips curves fail to beat a univariate benchmark model to certain states of the economy,

[3]Many studies focus on in-sample estimation, see e.g., Galí et al. (2001); O'Reilly and Whelan (2005); Doepke et al. (2008); Paloviita (2008); Musso et al. (2009); Montoya and Döhring (2011); Eser et al. (2020). There are less papers studying out-of-sample forecasting performance, see Rünstler (2002); Hubrich (2005); Canova (2007); Marcellino and Musso (2010); Buelens (2012); Bereau et al. (2018); Jarociński and Lenza (2018); Moretti et al. (2019).

namely expansion versus recession. Instead, the euro area inflation appeared to be particularly hard to forecast in the run-up to the EMU and in the aftermath of the sovereign debt crisis. In the first period inflation trended downwards and most models, being estimated on historical "high inflation" samples, resulted in overpredictions.[4] The "low inflation period" after the sovereign debt crisis was also particularly challenging for forecasters in terms of both capturing the inflation trend and the right measure of economic slack.

Second, in line with the US-based literature (see e.g., Clark and McCracken, 2010; Faust and Wright, 2013; Clark and Doh, 2014) we find that accounting for a trend or time-varying mean of inflation leads to sizable improvements in forecast accuracy in certain periods. As trend proxies we consider exponentially-weighted moving averages (EWMA) of past inflation rates, long-run inflation expectations available from Consensus Economics and lagged inflation (which results in the accelerationist version of the relationship). We find that removing a time-varying inflation trend helps, especially in the early part of the sample, corresponding to the inflation convergence period. The EWMA appears to be a good choice for estimating the inflation trend, as the gains with respect to the no detrending version prevail through most of the evaluation sample. Long-term measures of inflation expectations also bring forecast gains most of the time, but not over the latest low inflation episode.[5] The accelerationist version works well in the inflation convergence period and, interestingly, also during the latest low inflation period. The latter suggests that the persistence of inflation might have increased (see also Bonam et al., 2019). The former is in line with findings for the US (see e.g Peneva and Rudd, 2017; Ball and Mazumder, 2019a).

Third, regarding the relevant measure of slack we find that the output gap based on a simple filter performs relatively well, which is in line with the euro area analysis of Bereau et al. (2018) or Ciccarelli and Osbat (2017). We employ a wide set of product and labour market slack indicators, some of which have been constructed and back-casted for this investigation. In particular, we also explore a novel dataset of real-time estimates of the output and unemployment gaps published by international economic institutions (the OECD, the IMF, the European Commission and the Eurosystem/ECB). On the labour market it appears harder to single out one indicator that is superior over the entire period. In the last part of the sample a conventional filter-based unemployment gap fares worse than the broad measure of the unemployment rate (U6), which might better capture the structural changes that the labour market underwent after the Great Recession. The changing configuration of the labour market in terms of increasing informal jobs and other non-standard work agreements (see Bracha and Burke, 2016) might have rendered the estimation of labour market slack more problematic. While in general making use of the slack measures produced by international economic institutions does not provide for a 'silver bullet', they do seem to be quite useful in this particular period, after the double dip euro area recession, when deep structural economic changes rendered the estimation of slack more difficult.

Fourth, we do not find support for the inclusion of various external variables in order to enhance the out-of-sample predictive power of the Phillips curve. We link this result with the difficulty in forecasting such variables and admit that external supply-side shocks can have a considerable explanatory power for domestic inflation in-sample (see e.g. Bobeica and Jarociński, 2019, for evidence for the euro area).

[4] Fisher et al. (2002) argue that in times of monetary regime change model predictions based on economic activity might have no or only low explanatory power.

[5] Bańbura and van Vlodrop (2018) reach a similar conclusion using different models.

Fifth, new generation Phillips curve models (such as the one proposed by Chan et al., 2016, 2018) offer forecasting gains relative to a univariate benchmark and their main advantage appears to be related to the incorporation of time-varying trends and within-model estimation of economic slack.[6]

Finally, we find that the relative performance of different model variants changes over time. Forecasters should bear this in mind when betting on models which have performed well in the past. Pooling results from different models and averaging over certain modelling choices (e.g. across functional forms, estimation windows and lag selection criteria) and included variables offers some hedge against the instability in the forecast performance.

Our main conclusions are validated in a real-time forecast evaluation. We rely on a unique database for the euro area made available in the ECB's Statistical Data Warehouse. Due to the data availability, the real-time forecast evaluation is possible starting only in 2005. We cross-check the results regarding the institutional measures of slack, as these estimates are real time by nature and we confirm our main messages based on the pseudo real time analysis. We also conduct a real-time evaluation on the new generation Phillips curve models, as they appear to be a promising avenue for conducting forecast. We find that the real time forecast evaluation doesn't change the picture drawn based on pseudo real time data, with some models still offering consistent forecast gains with respect to univariate benchmarks.

Our paper is most closely related to Stock and Watson (1999), Stock and Watson (2009), Bereau et al. (2018) and Moretti et al. (2019). In contrast to the former two papers we focus on the euro area and also study a more recent sample. Compared to the latter two papers we consider a broader range of specifications over a longer period. We also look more in depth at the issue of time variation than all the papers mentioned. We refer to other related papers when detailing the modelling approaches and the results.

The paper is organised as follows. Section 2 presents the general setup of the forecast evaluation, the workhorse model and its forecasting performance. Section 3 evaluates whether enhancing the workhorse model along the dimensions listed above can offer gains in terms of forecast accuracy. Section 4 is the conclusion.

2 A workhorse Phillips curve model

2.1 Preliminaries

Before coming to the model let us provide some details of the forecast evaluations that follow.

Let $\pi_t = 400 \ln \left(\frac{P_t}{P_{t-1}}\right)$, where P_t is the appropriate (quarterly) price index, denote the annualised quarter-on-quarter inflation rate. Our target variable is the annualised h-period-ahead average inflation rate:

$$\pi_{t+h}^h = \frac{1}{h} \sum_{i=1}^{h} \pi_{t+i} = \frac{400}{h} \ln \left(\frac{P_{t+h}}{P_t}\right) . \tag{1}$$

[6] Canova (2007) shows that despite the evident structural changes in the process for inflation, the contribution of time variations in the coefficients to the forecasting performance of multivariate models is limited. Bańbura and van Vlodrop (2018) show that time varying trend is the key element that leads to better forecast performance for the VARs (as opposed to the coefficients or the variances).

P_t is the seasonally and working day adjusted euro area harmonized index of consumer prices (HICP) excluding energy. We disregard the energy component as it is more sensitive to (volatile) movements in oil prices and less to domestic real activity.[7]

Regarding the proxy for the real marginal cost or *slack* in the economy a wide range of choices has been considered in the literature. We focus the analysis on model versions that rely on "statistical" estimates of the *output gap* and the *unemployment gap*. The "statistical" gaps are obtained by applying the Christiano-Fitzgerald filter to logged GDP and to unemployment rate, where we keep the cycles shorter than 15 years.[8] Other slack measures, including composite indicators, are considered in Section 3.2. In addition, we evaluate slack estimates of international economic institutions in Section 3.2.2.

The baseline evaluations are conducted in a pseudo real-time fashion. This means that we disregard issues such as data revisions or publication delays (see e.g. Bańbura et al., 2013, for a discussion). In particular, at each point of the forecast evaluation sample, the filters are applied to a series ending at that point and based on the latest vintage of the data. We also conduct robustness checks over shorter samples and for selected exercises using a real-time database. In that case filters are applied to the appropriate vintage of the real activity measure, see below for details. The pseudo real-time estimates of the "statistical" slack measures are shown in Figure 12 in Appendix B. Figure 13 shows the slack estimates of international institutions.

All the data are quarterly and seasonally adjusted and cover the sample 1980 to 2018. The data details are provided in Appendix A.

We focus on one-year-ahead forecast horizon ($h = 4$).[9] The main evaluation criterion is the Root Mean Squared Forecast Error (RMSFE), but we also look at the continuous ranked probability score (CRPS) and log predictive scored (LPS) for the new generation models. The evaluation period is 1994-2018, unless indicated otherwise. In order to analyse how (relative) forecast accuracy has evolved over time we compute it over a rolling window of 20 quarters.

We compare the forecasting performance of the models relative to two *benchmarks* frequently employed in the literature: the "random walk" (RW) forecast of Atkeson and Ohanian (2001) where $\pi^h_{t+h|t} = \pi^4_t$ and the unobserved components stochastic volatility model (UCSV) of Stock and Watson (2007).[10].

[7]For the sake of robustness, we also looked at HICP inflation excluding energy and food; the overall messages are the same; the results are available upon request.

[8]Christiano-Fitzgerald filter is a nearly optimal one-sided band-pass filter, see Christiano and Fitzgerald (2003). Robustness checks were performed with the Hamilton (2018) approach and with the Hodrick-Prescott filter; also, robustness checks were performed whereby the underlying series were extended with AR process prior to detrending.

[9]Similar messages emerge also for longer horizons.

[10]The model is given by:

$$\begin{aligned} \pi_t &= \tau_t + \varepsilon_t, \\ \tau_t &= \psi_{t-1} + \eta_t. \end{aligned}$$

where ε_t and η_t are characterised by stochastic volatility. The forecast from this model is given by $\pi^h_{t+h|t} = \tau_{t|t}$. We use the noncentered parameterisation as in Chan (2018)

2.2 Forecasting performance

We start by investigating the out-of-sample forecast performance of a simple generic Phillips curve model, which we label as the "workhorse" model. It is an Autoregressive Distributed Lag (ADL) model specified as follows:

$$\tilde{\pi}_{t+1} = \alpha\tilde{\pi}_t + \beta y_{t+1} + \nu_{t+1} \tag{2}$$

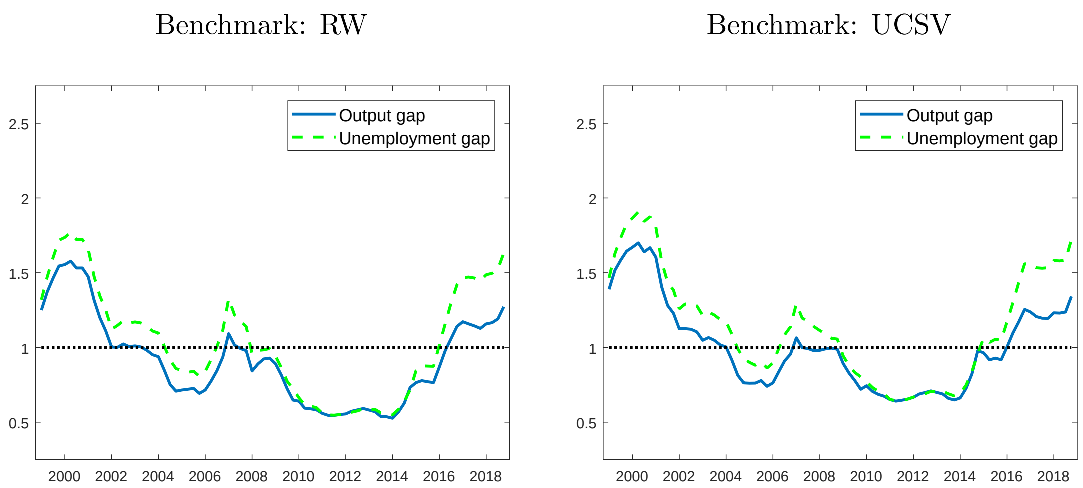
where α and β are the coefficients, y_t is a proxy for the real marginal costs (economic slack) and $\tilde{\pi}_t = \pi_t - \mu_\pi$ denotes the de-meaned inflation rate.[11] μ_π can be interpreted as an inflation trend that is constant over time. This assumption is relaxed in Section 3.1 where various forms for the trend process are evaluated.[12] The mean (or trend) is added back to the forecast.

The estimation is performed with a rolling window of 60 observations using the OLS.[13]

The forecasts h-steps-ahead are obtained in an *iterative* fashion. The explanatory variables are first forecasted with an AR(4) process.[14] Then we iteratively obtain forecasts for $\tilde{\pi}_{t+i}$, $i = 1, ..., h$.

Figure 1 illustrates the evolution of the forecast performance of two workhorse Phillips curve models including a slack measure for the product market, the output gap, and a slack measure for the labor market, the unemployment gap, relative to the two benchmarks, the RW and the UCSV.

Figure 1: Forecasting performance of the workhorse Phillips curve models



Note: RMSFEs over a rolling window of 20 quarters relative to the corresponding RMSFEs of the benchmark models.

The first result that stands out in Figure 1 is that the relative forecasting performance of the Phillips

[11]Such formulation, further augmented by a measure of inflation expectations and supply shocks proxy, was used to understand the drivers of inflation in the euro area by e.g. IMF (2013).

[12]Slack measures that are not conceptually mean 0 (as is the case for gaps) are also demeaned.

[13]Bayesian techniques yielded similar results.

[14]Dotsey et al. (2018) find that forecasts from Phillips curve models tend to be unconditionally inferior to those from univariate forecasting models, with some improvements brought by conditional forecasts. They pad future observations with forecasts from an AR(4) model for unemployment.

curve models changes considerably over time. The sample dependence of the Phillips curve performance was also documented for the US case by Fisher et al. (2002), Stock and Watson (2009) or (2010). Unlike Stock and Watson (2010), we do not find a link between the times in which Phillips curves fail to beat a univariate benchmark model and certain states of the economy, namely expansion versus recession. Instead, we find that for the euro area the (simple) Phillips curve models are outperformed by the univariate benchmarks before the inception of the EMU, before the financial crisis and more recently after the euro area sovereign debt crisis.[15] The Giacomini and Rossi (2010) fluctuation test indicates that the differences in forecasting performance are statistically significant only in the period before the inception of the EMU (see Figure 14 in Appendix C). These observations are valid for both measures of slack. Nevertheless, the output gap appears to perform marginally better in a rather consistent fashion and it is worth noting that the specification with the unemployment gap gets particularly worse towards the end of the sample. Finally, between the two chosen benchmarks, the UCSV yields inflation forecasts which are very similar, but just marginally better than that of the random walk, so we will use only the UCSV as the benchmark in what follows.

In order to get further insights into the forecast performance, Figure 2 plots the model forecasts against the outcomes. The periods when the performance of the workhorse Phillips curve models worsens compared to univariate benchmarks are characterised by fundamentally different economic environments. The first episode corresponds to the run up to the euro introduction when the inflation rates in many euro area countries converged to lower levels, also given the anchoring of inflation expectations with the new monetary authority mandated to ensure price stability. Over this period the workhorse Phillips curve models, which assume a constant mean in the inflation rate, failed to capture the pronounced downward trend in inflation and yielded a systematic overprediction of the inflation rate. Inflation appears harder to forecast again towards the end of the sample, with model predictions systematically higher than the outcomes. The issue appears the most severe in the version relying on the unemployment gap as the measure of slack. That is related to the strong improvements in the euro area labour market in the recent years (at least when looking solely at the headline unemployment rate indicator), which makes measures based on statistical filters (as those applied here) to point to rather buoyant labour market conditions (see Figure 12 in Appendix B) at a time when inflation was stubbornly low.

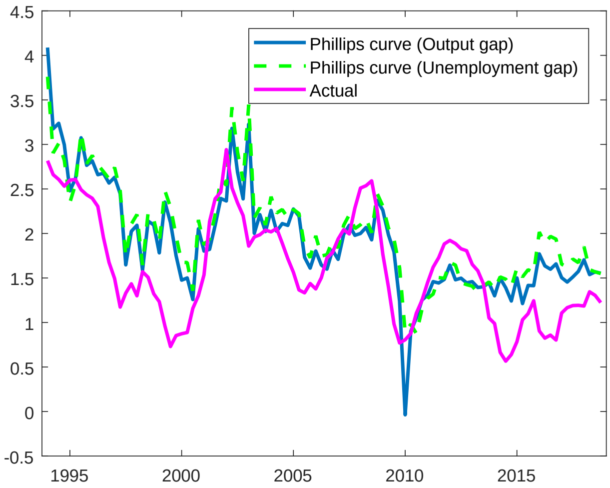
There has been an intense discussion surrounding the two puzzling inflation episodes in the aftermath of the Great Recession, namely the 'missing disinflation' and the subsequent 'missing inflation' (see Bobeica and Jarociński, 2019, for a comprehensive analysis of these two puzzles for the euro area). It appears that for the euro area and the class of models considered here the second episode, the 'missing inflation' starting in 2013 is raising more challenges.

2.3 Robustness

The question arises whether the lessons derived above are driven by specific modelling choices or hold across a wider range of Phillips curve specifications. This is an important aspect to be considered, as

[15] See also Figure 27 in the Appendix, which uses a narrower rolling window and shows e.g. a marked worsening in relative forecast performance during the sovereign debt crisis; similar patterns can be observed based on the absolute forecast errors (lower panels of the same figure). The findings also hold for the versions allowing for a time-varying trend.

Figure 2: Actual inflation versus Phillips curve forecasts



when it comes to specifying a Phillips curve model, the model uncertainty is sizable. In this subsection we assess the robustness of the results across the following choices:

1. Estimation window: we also apply a rolling window of 40 quarters (in addition to 60 used above) and *recursive* estimation with the estimation sample starting in 1980.[16]

2. Number of lags included and lag selection method: we allow up to four lags of inflation and of slack in Equation (2) and at each step of the evaluation we choose the number by either the AIC or the BIC criterion.

3. "Functional" form of the model: in addition to the one given by Equation (2) we consider the following specifications:

   - *ADL model with "direct forecast" formulation*:

   $$\tilde{\pi}^h_{t+h} = \alpha(L)\tilde{\pi}_t + \beta(L)y_t + \nu^h_{t+h} \tag{3}$$

   where $\alpha(L)$ and $\beta(L)$ are lag polynomials and h is the forecast horizon, i.e. 4 quarters. Such "direct forecast" specifications have been often employed in forecasting applications following Stock and Watson (1999) and Stock and Watson (2009).[17]

   - *VAR model*:

   $$X_{t+1} = \Phi(L)X_t + \nu_{t+1}\,, \qquad X_t = [\tilde{\pi}_t \; y_t]' \tag{4}$$

[16] Rossi and Inoue (2012) discuss the issues related to the choice of the estimation window in forecast evaluations and recommend comparisons over a wide range of window sizes. Pesaran and Timmermann (2007) recommend combining forecasts based on different window sizes in order to deal with structural breaks at unknown points in time.

[17] Note that this specification is different from Stock and Watson's in that they impose that inflation process is integrated of order one ($\alpha(1) = 1$). We discuss and evaluate such specifications in Section 3.1.

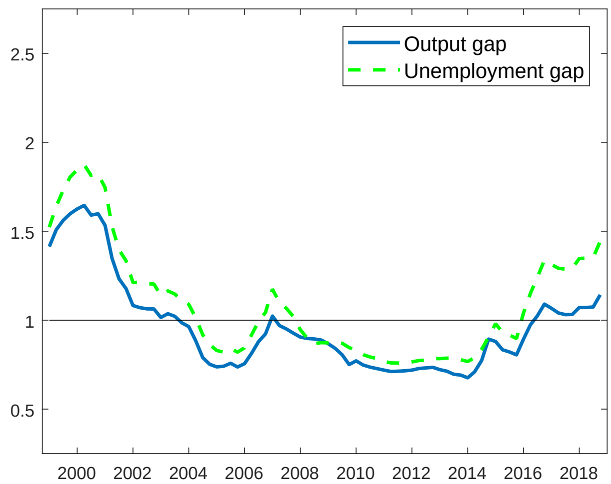
"Phillips curve" VARs were used by e.g. Garratt et al. (2014) and Hubrich (2005) to forecast inflation.

- *ADL model with a "lagged" slack measure*:

  We replace y_{t+1} by y_t in Equation (2). This particular functional form is motivated by the findings of previous work on the euro area, see e.g. Ciccarelli and Osbat (2017) and Bobeica and Sokol (2019).

The combination of model versions described above results in 36 specifications for each slack measure. Figure 3 presents the relative performance of the average forecast (across the specifications for each slack measure). The messages remain qualitatively the same compared to those from the workhorse model in Figure 1. It appears, nevertheless, that some accuracy gains can be attained by forecast averaging as the accuracy of the average across the specifications is higher than that of the workhorse model, in particular in the periods when inflation is harder to forecast (relative to the benchmark model). Interestingly, no particular specification choice among the ones listed above outperforms the others in a systematic fashion, see Figure 15 in Appendix C, which compares accuracy along one "dimension" of forecasts averaged along the remaining specification choices.

Figure 3: Forecasting performance of average over different specifications



Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model.

Another issue is how robust the results are to choosing another univariate benchmark. Figure 17 in Appendix C reports the RMSFEs of the Phillips curves relative to their univariate counterparts, i.e. models specified by a version of Equation 2 in which the slope coefficient β is restricted at 0 (for the forecasts averaged over all considered functional forms, estimation windows and lag selection approaches as discussed in this section). The performance of the Phillips curves including the output gap is better than that of the univariate version for most of the sample. This indicates that including output gap helps in forecasting inflation even if the improvements are modest. The gains of including an unemployment gap are not visible all the time, with a deterioration especially towards the end of the sample, as previously discussed.

3 Can one improve on the workhorse Phillips curve model?

The previous section documented a weak forecast performance of the simple Phillips curve model in some periods and hinted at possible underlying reasons such as changing inflation trend or difficulties in capturing the degree of slack in the economy. In this section we investigate whether alternative specifications of these and other ingredients of the model result in better forecast performance. In more detail, we conduct comprehensive comparisons along the following dimensions: inflation trend specifications, measures of economic slack (including slack estimates from international economic institutions), additional controls in the shape of external/supply shock variables, further features as incorporated in more recent ('new generation') specifications of Phillips curve models (most notably general forms of time variation). Given the results in Section 2.3 we report the accuracy of average forecast across all the specifications considered in that section (different estimation windows, lag selection approaches and functional forms), unless indicated otherwise.[18]

3.1 Inflation trend

The Phillips curve models presented so far assumed a constant (unconditional) mean in inflation over the estimation sample. Yet, many papers found that there are forecast gains when assuming a time-varying inflation trend (see e.g. Clark and McCracken, 2008; Faust and Wright, 2013; Zaman, 2013; Clark and Doh, 2014; Bańbura and van Vlodrop, 2018). The inflation trend is an unobservable variable, surrounded by considerable model uncertainty. Clark and Doh (2014) show that the relative forecast performance of different (trend) specifications varies over time. However, they also find that a model capturing the trend with the long-term inflation expectations from surveys is consistently among the best models, as is a local level model. We evaluate different approaches considered in the literature, we use "statistical" trend estimates based on past inflation rates and we also rely on long-term survey expectations of inflation as proxies. Precisely, we take $\tilde{\pi}_t = \pi_t - \pi_t^{TR}$, with the following specifications of π_t^{TR}:

- *Previous period inflation rate (accelerationist)*: $\pi_t^{TR} = \pi_{t-1}$

  This approach amounts to estimating the models with inflation in differences for the ADL indirect forecast and the VAR versions: $\tilde{\pi}_t = \Delta\pi_t$. For the ADL direct forecast we use the formulation of Stock and Watson (1999, 2009).[19] This is the *accelerationist* Phillips curve (Friedman, 1968), where gaps are related to the *changes* as opposed to the *level* of inflation.

[18] Results remain qualitatively the same when we consider the specifications as in the workhorse model.

[19] In order to have a specification as in Stock and Watson (1999, 2009) we also include an intercept in Equations (2) - (4). We also consider versions without an intercept as e.g. the "triangle" model of Gordon (1982; 1990). Forecast performance of both versions is compared in Figure 16 in the Appendix. The latter version performs somewhat better than the former, however, relatively to other measures of trend the results are qualitatively similar.

- *Exponentially-weighted moving average (EWMA) of past inflation*: $\pi_t^{TR} = \phi \sum_{j=0}^{\infty}(1-\phi)^j \pi_{t-j}$

  This approach assumes a "statistical" trend, based on weighted average of past inflation rates with a "smoothing" parameter ϕ. Clark and McCracken (2008, 2010) document improved forecast performance over some periods for VARs in which inflation is detrended using the EWMA compared to non-detrended versions. We choose a fixed ϕ equal to 0.05.[20]

- *Long-term survey inflation expectations*: $\pi_t^{TR} = \pi_t^{Cons} - (\mu_{Cons} - \mu_\pi)$

  π_t^{Cons} is the forecast of average inflation 6 to 10 years ahead from Consensus Economics. The second term is the difference of historical means of the expectations and of the target variable and corrects for the fact that the Consensus forecast concerns headline inflation and inflation excluding energy has been systematically lower over the sample considered (bias correction).[21]

  Long-term survey-based inflation expectations have been often used to proxy trend inflation (see e.g. Faust and Wright, 2013; Yellen, 2015) and they might be better suited than model- or filter-based trends to account for expected changes in policies, such as those adopted during the run-up to the introduction of the euro. In terms of forecasting, studies have found improvements in the forecast performance by modeling the inflation gap (as opposed to inflation itself) as a deviation from a survey based measure of inflation expectation (see e.g. Faust and Wright, 2013; Zaman, 2013; Clark and Doh, 2014). Our formulation is similar to that recently used by Hooper et al. (2019).[22] The approach is also related to the approach of Wright (2013), who uses inflation expectations as priors for the mean of inflation in a VAR and related to the shifting endpoint concept of Kozicki and Tinsley (2001).

The pseudo real-time estimates of the trends following the three approaches are presented in Figure 11 in Appendix B.[23] Notably, survey expectations declined more rapidly in the run-up to the introduction of the euro compared to the EWMA estimates, while not being more noisy. This suggests that indeed the forecasters were quicker in changing their beliefs about trend inflation in light of the new economic environment than what might have been inferred based solely on filtered past inflation data (see Faust and Wright, 2013, for similar observations for the US). Towards the end of the sample, trend estimates based on filtered inflation are sizably lower than those based on survey expectations, with the latter remaining relatively stable. Finally, previous period inflation rate is a very noisy measure of

[20] Clark and McCracken (2008, 2010) use the smoothing parameter of 0.05 or 0.07. We have also considered ϕ equal to 0.15, 0.25 or estimated from the first-order integrated moving average (IMA(1,1)) model. Regarding the latter, note that π_t^{TR} would be the forecast for the IMA(1,1) process with $\phi = 1 + \psi$, where ψ is the moving average coefficient. Further IMA(1,1) is equivalent to the UCSV model with a constant ratio of variances of temporary and permanent shocks (see Stock and Watson, 2007). These choices did not result in consistently better forecast performance, see Figure 16 in the Appendix. In particular, the performance of IMA(1,1) model is very similar to the benchmark UCSV model, indicating that allowing for time variation in the relative variances of temporary and permanent shocks does not lead to improvements in forecast accuracy.

[21] The bias corrected version of the specification yields better results over the recent sample, see Figure 16 in the Appendix.

[22] These authors include the long-term inflation expectations as a regressor, however they impose that the sum of the coefficients on lagged inflation and on the expectations is equal to 1. If we abstract from the intercept and "other factors" which they also include in the equation and if the long-term expectations do not vary strongly over time (which is usually the case) then the two formulations are equivalent.

[23] The pseudo real-time estimates of trend based on Consensus forecasts change over time as the means in the bias correction are calculated in a pseudo real-time fashion.

trend. Over the forecast horizon the trends are assumed to remain constant for the EWMA and survey expectations approach but are model consistent for the accelerationist approach.

Figure 4 compares the relative forecast performance under different assumptions for trend inflation for the output and unemployment gap measures considered above. As long-term inflation expectations from Consensus Economics are available only as of 1990, the evaluation sample for this approach starts in 2000 (instead of 1993).

Whereas the relative forecast performance of different detrending methods changes over time, we find that detrending is helpful, particularly in the early part of the sample, corresponding to the inflation convergence period. Detrending with survey inflation expectations and with EWMA helps for most of the analysed period. The performance related to trends based on surveys, while favourable compared to other methods over most of the sample, markedly deteriorates over the last part (when they were quite upbeat despite a protracted period of low inflation). By contrast, the EWMA trend improves on the models without detrending also towards the end of the sample. This is related to the differences in behaviour of the two trend estimates in the period of low inflation discussed above.

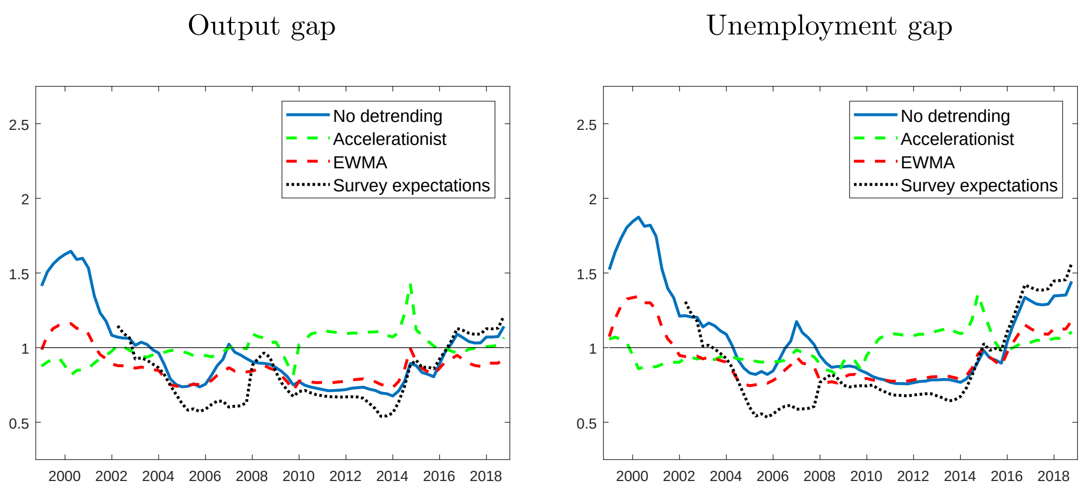
An interesting result refers to the performance of the accelerationist Phillips curve models. They perform the best in the 90s and early 2000. A similar story was found for the US, where the accelerationist version appears to fit the data starting in the 60s, but fails to characterize inflation in the last couple of decades. Ball and Mazumder (2019a) show that the change in the relationship between inflation and unemployment from an accelerationist Phillips curve to a level-level relationship is due to the changing behaviour of inflation expectations, which turned from backward-looking until the late 1990s to firmly anchored by the Fed's inflation target. In a similar vein, Peneva and Rudd (2017) argue that fitting an accelerationist specification when inflation is characterised by a long-run stable trend is challenging. In the case of the euro area the accelerationist models have witnessed some improvement in their forecasting power also towards the end of the sample. This might suggest that inflation persistence has increased during the low inflation period in the euro area, which is in line with the findings of Ciccarelli and Osbat (2017).

Figure 17 in the Appendix compares the performance of the non-detrended and detrended versions to the univariate counterparts (without a slack measure). Also in case of detrended versions adding the output gap helps in improving forecast performance. In addition, Figure 18 reports the performance of the Phllips curve models relative to a "pure trend forecast" showing that including the lagged terms of the inflation gap (the difference between inflation and the trend) and a slack measure leads to sizable improvements in forecast performance for most specifications over most of the sample.

3.2 Measures of economic slack

How to measure the degree of economic slack or labour market tightness is a key consideration when specifying a Phillips curve model. After the Great Recession, the discussion related to the well known difficulty in constructing real-time estimates of the output gap (Orphanides and van Norden, 2005) intensified, given the ambiguous impact of the crisis on potential growth and more generally the secular stagnation debate (see e.g. Blanchard et al., 2015, or Jarociński and Lenza, 2018). In addition, the challenge in reconciling nominal developments with traditional indicators of economic or labour market slack made economists believe that perhaps alternative measures were needed (see e.g. Yellen, 2019).

Figure 4: The impact of inflation detrending on the forecasting performance



Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model.

In this section we evaluate a wide range of measures in terms of usefulness for forecasting euro area inflation. We distinguish between "statistical" measures based on simple transformations, filtering or econometric models and "institutional" measures that rely on richer data and modelling frameworks as well as on judgment.

3.2.1 Statistical measures of slack

We divide the measures into the following three groups:

- *Product market slack measures*

  We consider both hard indicators, as well as survey based measures, as follows: (i) output gap; (ii) GDP growth; (iii) industrial production growth; (iv) industrial production gap; (v) capacity utilisation; (vi) economic sentiment.

- *Labour market slack measures*

  We consider conventional indicators, as well as measures which were proposed more recently in the literature to better grasp the degree of labour underutilization: (i) unemployment gap; (ii) unemployment rate; (iii) the unemployment recession gap of Stock and Watson (2010); (iv) employment growth; (v) employment gap; (vi) short-term unemployment rate; (vii) broad unemployment rate (U6) and (viii) U6 unemployment gap.

  The *short-term unemployment rate* was brought forward in the US case as an explanation for the 'missing deflation' episode, when inflation did not fall as much as the headline unemployment rate would have predicted (see Gordon, 2013, or Ball and Mazumder, 2019a). At that time the short-term unemployment rate increased much less than total unemployment and there are reasons to believe that this particular indicator is more indicative for inflationary pressures: those unemployed for longer are less attached to the labour force, search less intensively for work and are less attractive to employers. *Broader measures of unemployment rate*, such as the so-called U6 rate, were brought into the picture later on, during the recovery after the Great Recession, mainly in an attempt to explain the 'missing inflation' episode. The headline unemployment rate was improving fast, but at the same time the labour market posted more underemployed and more people marginally attached to the labour force, which is captured in broader measures of labour underutilisation (see the discussion in Coeure, 2017 for the euro area and Bell and Blanchflower, 2018 for Europe and the US). These two indicators have not previously been considered in longer term evaluation for the euro area due to their short history. For the purpose of the analysis here they have been back-casted (see details in Appendices A and B).

- *Composite slack measures*

  In order to synthesise a broader information set, we also consider two composite indicators, namely (i) the first principal component of all variables listed above and (ii) an output gap measure enhanced with financial variables.

  The former has been shown to perform better than individual measures by Stock and Watson (1999). The latter encompasses the idea that some financial and asset price indicators can reflect and even cause business cycle fluctuations, as the experience of the Great Recession shows. The measure considered here follows the methodology proposed in Melolinna and Tóth (2018), which is very similar to Borio et al. (2015), but instead of considering only real credit growth and residential property price indices, a financial conditions index is used, comprising: the real growth rate of credit to households, real growth rate of credit to non-financial corporations (NFCs), real growth rate of a broad monetary aggregate (M3), real growth rate of residential property prices and the spread between short- and long-term risk free interest rate.

All "gap" measures (apart from the last one) are obtained using the Christiano-Fitzgerald filter.

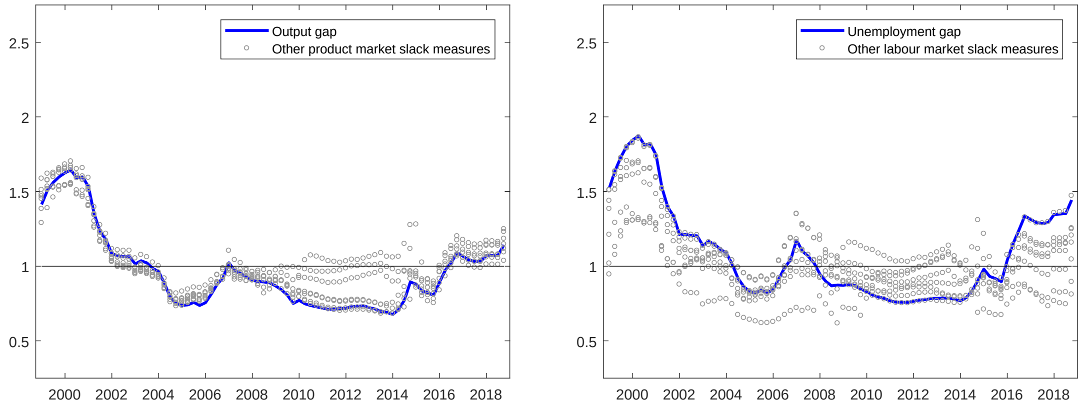
Figure 5 shows the forecasting performance of Phillips curve models with different slack measures (average across specifications listed in Section 2.3) for model versions without and with EWMA detrending. More detailed information is provided in Figure 19 in Appendix C. For the product market, the benchmark (filtered based) output gap measure fares well compared to alternatives for most of the sample. For the labour market other indicators, in particular the broad measure of the unemployment rate, U6, or the headline unemployment rate, seem to work better than the benchmark unemployment gap measure in the last part of the sample. In contrast to the results in Stock and Watson (2010), the unemployment recession gap does not seem to bring gains compared to other indicators. Turning to the synthetic measures, the output gap enhanced with financial variables does not appear to bring additional predictive power in the post Great Recession period. The principal component works fine, but generally worse than the output gap.

All in all, the choice of the slack measure matters, with some bets safer than the others. The output gap fares well compared to alternatives. On the labour market it is harder to pin down an indicator which performs well across the entire period. Especially in the last part of the sample, alternative indicators to the unemployment gap bring some forecast gains, perhaps related to structural changes in

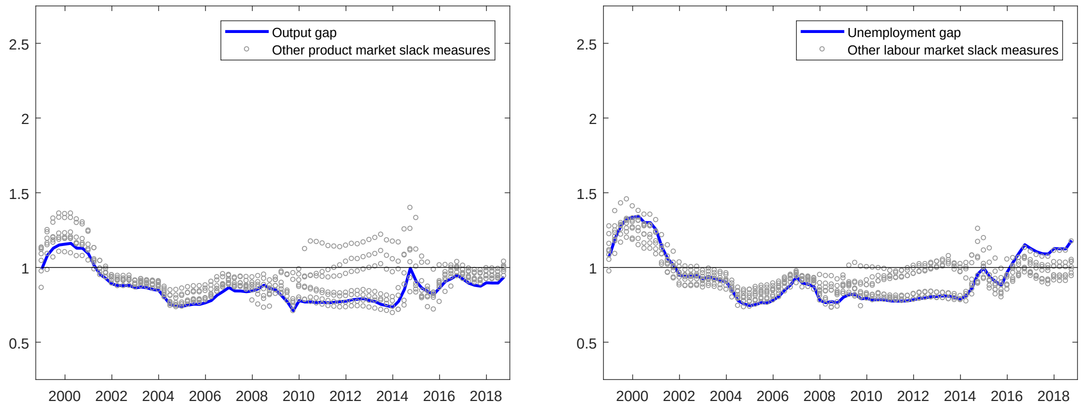
Figure 5: The impact of slack measure choice on the forecasting performance

Slack in product market　　　　Slack in labour market

No detrending



Inflation detrending with EWMA



Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. '*' represent the RMSFEs of models with other measures of slack from a given group. The principal component has been included in both panels; the output gap with financial variables has been included in the panel with slack in the product market.

the labour market in the aftermath of the Great Recession not captured by this conventional indicator. These results talk in favor of looking at a broader set of indicators when assessing the labour market slack and its implications for inflation. Finally, for all classes of indicators detrending helps.

3.2.2 Measures of slack from international economic institutions

In this section we investigate whether more sophisticated methods to estimate economic slack than the ones used so far lead to more accurate inflation forecasts. For this we exploit the estimates by international economic institutions, namely the OECD, the IMF, the European Commission (EC) and the Eurosystem/ECB[24]. These institutions pay close attention to the degree of slack in the product and labour market, as these are key macroeconomic indicators for forecasting and policy design. They have implications for the dynamics of wages and prices, but also for the accumulation of imbalances, for the analysis of the cyclically adjusted budget balances or the sustainability of public debt. A macroeconomic production function supplemented with filtering and other econometric methods is usually applied by these institutions to estimate the output and unemployment gaps, although judgement is also at play (see e.g. Andersson et al., 2018).

We exploit a novel data set of real-time vintages of slack estimates from these institutions:

- *OECD:*
  Output gap: vintages starting with the spring 2000 OECD Economic Outlook; unemployment gap: vintages starting the fall 2001 OECD Economic Outlook.
- *IMF:*
  Output gap and unemployment gap: vintages starting with the October 2004 World Economic Outlook projections.
- *EC:*
  Output gap: vintages starting with the fall 2002 economic forecast; unemployment gap: vintages starting with the spring 2003 economic forecast.
- *Eurosystem/ECB*, labeled *Eurosystem*:
  Output gap and unemployment gap: vintages starting with the December 2008 Broad Macroeconomic Projection Exercise.

As the available vintages do not cover the entire evaluated period (1994-2018), for this particular exercise the evaluation sample is much shorter.[25] The slack measures are usually updated and made available with the official forecasts. The OECD, the IMF and the European Commission thus only provide annual estimates two times per year. We interpolate the series to quarterly frequency using the Stram-Wei method and in order to produce forecasts at quarterly intervals we assume that in between the updates the estimates of the gaps are unchanged (Figure 13 in Appendix B shows how these slack measures look like). In the pseudo real-time forecast evaluations we apply the following "timing" assumption: measures of slack released in the spring forecasts by the OECD and EC, in April by IMF and in March by the Eurosystem/ECB incorporate actual data up to the fourth quarter of the

[24] These refer to the (confidential) estimates of output and unemployment gaps prepared by staff in the context of the Eurosystem/ECB macroeconomic projections, see ECB (2016).

[25] In addition, the available time series for slack are shorter - in the case of the OECD and the IMF, estimates tend to start in '91, in the case of the EC and the Eurosystem/ECB, the series tend to start in '95.

previous year.[26]

Figure 6 compares the relative forecasting performance of the Phillips curve models relying on slack measures provided by the international institutions to those incorporating the "statistical" output and unemployment gaps. Again versions without and with EWMA inflation detrending are considered. Up to the most recent recovery, a forecaster trying to predict inflation using the Phillips curve models considered here and taking slack estimates from international economic institutions would not have been better off than applying a simple statistical filter to derive the output or the unemployment gap. Interestingly, the relative performance of "statistical" and "institutional" slack estimates changes in the recent years. This relates to the fact that compared to the former, the latter gap measures are less buoyant towards the end of the sample, which makes them more in line with inflation developments. In particular, they imply less tight labour market conditions. As a result, the models with slack estimates from the institutions consistently outperform the UCSV benchmark during the economic recovery after the Great Recession. The results tend to hold irrespective of whether inflation is detrended or not and across institutions, as the estimates tend to be similar across them. The Eurosystem/ECB measures appear to bring some small comparative gains. Based on this one might infer that when the uncertainty regarding slack is particularly high, it pays off to use information from the international economic institutions. These economic institutions also produce forecasts for the slack measures so one might be tempted to produce conditional forecasts taking this information into account. The results of such an exercise are presented in Figure 20 in Appendix C.[27] Overall, the improvements to our approach where we extend the slack measures via an autoregressive model are marginal and they mainly refer to the period when the performance is particularly poor.

A fair point that one can raise regarding the comparison in Figure 6 is that the "institutional" slack estimates are truly real-time, i.e. based on the data available for the respective forecast exercise, whereas the "statistical" slack indicators are based on the latest available release of real GDP and unemployment rate, which can play to their advantage.

To check the robustness of the results along this dimension we perform an analogous exercise as reported in Figure 6 but based on real-time data for all the variables. For this purpose we construct real-time vintages for inflation, real GDP and unemployment rate starting in mid 2005, using the information stored in the ECB's Statistical Data Warehouse (SDW).[28] The cut-off dates for the real-time vintages are set so that they approximately match those for the data used for the respective

[26] For example, in order to forecast the average inflation rate over 2010Q1-2010Q4 we use the output/unemployment gap estimate up to 2009Q4 from the spring 2010 forecast for the OECD and EC, April 2010 forecast for the IMF and March 2010 Macroeconomic Projection Exercise for the Eurosystem/ECB. For the subsequent period, in order to forecast the average inflation rate over 2010Q2-2010Q3 we use the output/unemployment gap estimate up to 2010Q1 from the aforementioned projection rounds for the OECD, EC and IMF (as these three institutions publish the results of their forecast exercise only twice per year) and from the June 2010 Broad Macroeconomic Projection Exercise of the Eurosystem/ECB, which publishes projections every quarter.

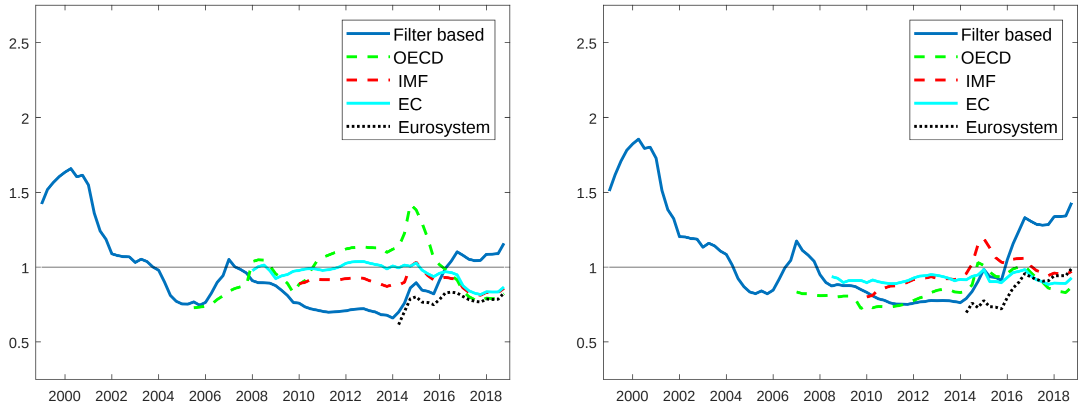
[27] As conditioning is not possible for the *ADL direct* functional form, it is not included in the set of specifications for this exercise.

[28] Earlier vintages have not been systematically stored in the SDW. For some variables they are available from the database constructed by Giannone et al. (2012) but e.g. not for seasonally adjusted HICP excluding energy. For GDP the first two vintages are pseudo real-time as the respective real-time data is not available from the SDW.

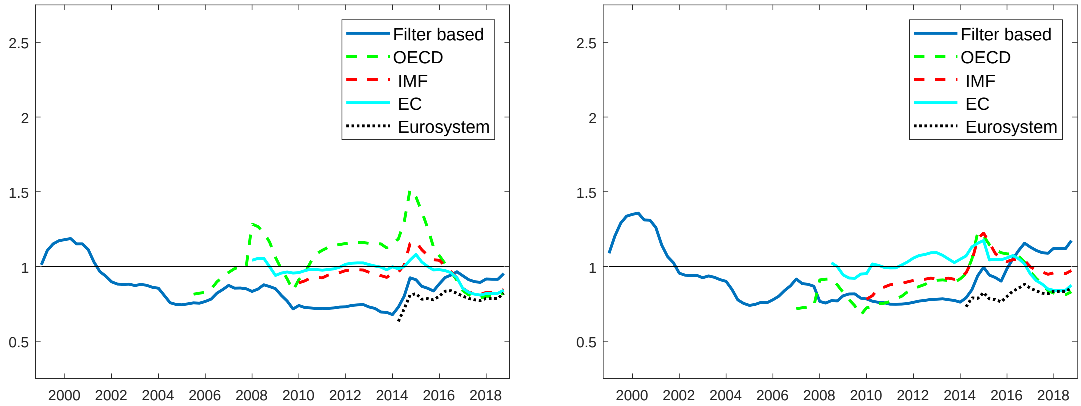
Figure 6: Forecast performance with slack estimates from international institutions

Output gap | Unemployment gap

No detrending



Inflation detrending with EWMA



Note: Average RMSFEs over ADL 1Q and Workhorse model specifications, but also across estimation samples and lag choice approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model.

forecasts by the international institutions.[29] Figure 21 in Appendix C shows the results. The messages drawn so far continue to hold. The "statistical" measures of slack perform well when compared to the "institutional" ones (with Eurosystem/ECB measures remaining in the "lead") with the exception of the unemployment gap in the recent period. Also, inflation detrending helps irrespective of the chosen measure of slack, "statistical" or "institutional".

[29] More precisely, we apply the following "timing" assumption: measures of slack released in the spring forecasts by the OECD and EC and in April by IMF incorporate actual data as of 25th of January of that year, whereas in the case of the March projections by the Eurosystem/ECB the cut-off is 25th of February.

3.3 Cost-push shocks and external developments

Ever since the development of the "triangle model" of inflation (Gordon, 1982, 1990) cost-push shocks, particularly coming from non-domestic drivers, have been considered to be relevant for domestic prices. In-sample, there is some evidence that external developments, as reflected in commodity prices, exchange rates, foreign slack or inflation, can account for part of domestic inflation, arguably more so for total inflation than for a "core" index stripped out of volatile items (see e.g. Bobeica and Jarociński, 2019; Forbes et al., 2019; Forbes, 2019). Nevertheless, when it comes to out-of-sample forecasting, the usefulness of such indicators can be undermined by the fact that they are volatile and/or difficult to forecast and the relationships might be time-varying (see Bereau et al., 2018).[30]

We consider the following measures of external developments:

- *Import, producer and foreign prices:*

  (i) the total import deflator; (ii) the extra euro area import deflator; (iii) the producer price index (PPI) for total industry less construction and the (iv) US CPI.

- *Commodity prices:*

  (i) the oil price in euros and a (ii) non-energy commodity price index in euros.

- *Exchange rates*

  (i) the nominal effective exchange rate of the euro and (ii) the bilateral euro dollar exchange rate.

The choice of variables is also motivated by the availability of long historical data. All variables are transformed to annualized quarter-on-quarter growth rates. We augment Equation (2) and the versions presented in Section 2.3 with one "external" variable at a time.

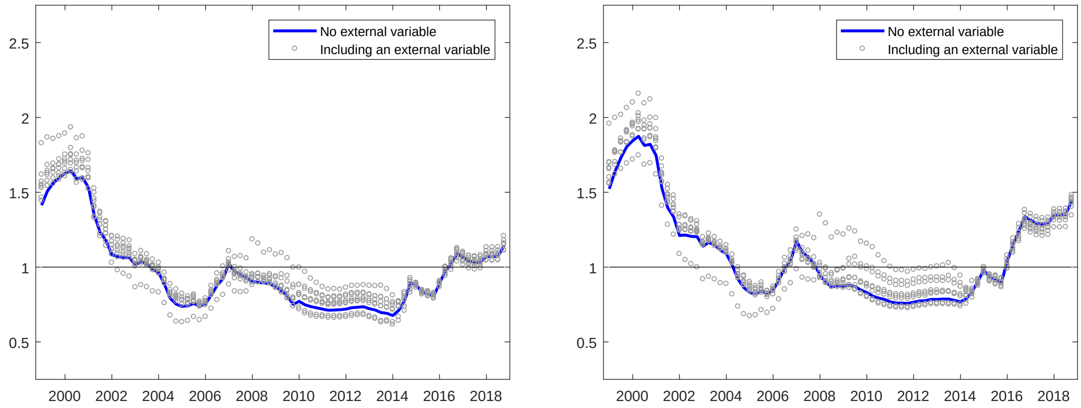
Figure 7 compares the forecasting performance of Phillips curve models augmented with different external variables to the versions with only a slack measure as an explanatory variable (average across model versions considered in Section 2.3 for the two benchmark slack measures). Versions without and with EWMA detrending are considered. Some limited forecast gains are entailed in mid-2000s in non-detrended version by including US inflation (see Figure 22 in the Appendix). Interestingly, this acts more as a trend substitute, as the forecast gains disappear in specifications accounting for a time-varying inflation trend. Otherwise, including external variables does not enhance the predictive power of the Phillips curve model suite for HICP inflation excluding energy. Similar messages emerge when looking at average forecast across slack measures considered in Section 3.2, see Figure 23 in the Appendix. We confirm thus the result obtained by Bereau et al. (2018) for the euro area and we interpret this as being related with the difficulty in forecasting these variables. This does not preclude that they might have in-sample explanatory power in certain episodes.

[30] IMF (2013) shows that the importance of the import price inflation for total inflation has increased since the year 2000. Similarly, Forbes (2019) argues that the explanatory power for total inflation of external developments, particularly of non-fuel commodity prices, has increased since the financial crisis.

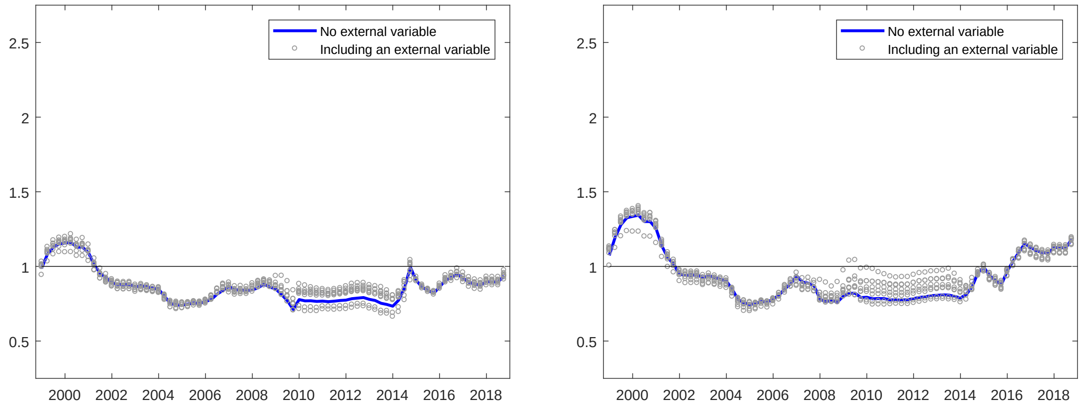
Figure 7: The impact of considering external developments on the forecasting performance

Output gap | Unemployment gap

No detrending



Inflation detrending with EWMA



Note: Average RMSFEs over all considered functional forms, estimation windows and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. '*' represent the RMSFEs of models which include an external variable.

3.4 New generation Phillips curve models

In the specifications considered so far, the only type of time variation that was allowed for was captured by the inflation trend. However, several studies have postulated that other types of time-varying features in the Phillips curve relationship might be relevant, most notably what regards the slope. In particular, the puzzling behaviour of euro area inflation since the Great Recession has generated an intense debate on whether the Phillips curve has flattened or steepened (see e.g. Constancio, 2015; Riggi and Venditti, 2015; Ciccarelli and Osbat, 2017).[31] While some studies reconcile in-sample inflation dynamics with a measure of slack by resorting to time-varying parameter models, the forecasting properties of such models for the euro area have not been thoroughly analysed.

We investigate whether some recently proposed approaches that allow for time variation in the Phillips curve bring forecast gains relative to more traditional models in the case of the euro area. We narrow down the analysis by focusing on features such as time variation in the coefficients, stochastic volatility in the residuals, explicit modelling of an inflation trend and an endogenously estimated slack measure. We consider the following models:

- *A (general) time-varying parameter Phillips curve (TVP SV)*

$$\pi_{t+1} = c_{t+1} + \alpha_{t+1}\pi_t + \beta_{t+1}y_{t+1} + \nu_{t+1} \tag{5}$$

  The Phillips curve coefficients and log volatility of the residuals follow random walks. Inflation trend is not explicitly modelled, rather a (time-varying) intercept is included in the equation. Similar models were used in e.g. Fuhrer et al. (2009), Benkovskis et al. (2011) or Groen et al. (2013). This is a single equation equivalent of the time-varying parameter VAR model proposed by Primiceri (2005). Fuhrer et al. (2009) and D'Agostino et al. (2013) find that introducing such type of time variation helps when forecasting inflation.

- *A Phillips curve model á la Chan et al. (2016)* (henceforth *CKP16*)

$$(\pi_{t+1} - \pi_{t+1}^{TR}) = \alpha_{t+1}(\pi_t - \pi_t^{TR}) + \beta_{t+1}(u_{t+1} - u_{t+1}^{\star}) + \nu_{t+1} \tag{6}$$

  where $u_t^{\star}$ is an endogenous measure of natural rate of unemployment (in other words, the measure of slack, $u_t - u_t^{\star}$, is estimated within the model). Inflation trend and the natural rate of unemployment evolve as bounded random walks. On top of the theoretical appeal, from a technical point of view, the imposed bounds can be another source of information that lessen the need for tight priors or other restrictions imposed on the model.[32] The coefficients and log volatility of the residuals follow random walks. Similar models have been estimated for the euro area inflation by Hindrayanto et al. (2019)[33] or by Stevens and Wauters (2018)[34].

- *A Phillips curve model á la Chan et al. (2018)* (henceforth *CCK18*)

$$(\pi_{t+1} - \pi_{t+1}^{TR}) = \alpha_{t+1}(\pi_t - \pi_t^{TR}) + \beta_{t+1}y_{t+1} + \nu_{t+1} \tag{7}$$
$$\pi_{t+1}^{Cons} = a_{t+1} + b_{t+1}\pi_{t+1}^{TR} + u_{t+1} \tag{8}$$

[31] Earlier evidence on time variation in the Phillips in the euro area can be found e.g. in Musso et al. (2009) and Benkovskis et al. (2011). Musso et al. (2009) find support for time variation in the mean and the slope of the euro area Phillips curve and propose to employ a smooth transition model. We do not consider this approach here.

[32] Such models can be difficult to estimate without restrictions or strong prior information, as the likelihood function can become flat on several dimensions, see Stella and Stock (2015), who build a similar model but without the bounds on the random walk processes.

[33] They propose a more refined process for the unemployment gap and do not impose boundedness on the random walks. They estimate the model by maximum likelihood.

[34] The authors also include external factors and inflation expectations with various horizons; they show that the short- and medium-term horizon expectations bring useful information for forecasting inflation.

The inflation trend follows a random walk but it is also linked to the long-term inflation expectations via a measurement equation. The appealing feature of the model is that it doesn't simply equate long-term forecasts from surveys with inflation trends, as surveys can be biased[35] and are available for only a limited range of inflation measures. Instead, it allows for a potentially time-varying linear relationship between trend inflation and the long-term expectations subject to a (measurement) error. The Phillips curve coefficients and log volatility of the residuals follow random walks.

- *A Phillips curve model á la Chan et al. (2018) without linking the inflation trend to inflation expectations* (henceforth *CCK18 no Exp*)

  This is a version of the previous model where the second (measurement) equation is not included.

Estimation is carried out in a Bayesian setting, in a recursive fashion starting in the '80s.[36] As *CCK18* relies on long-term inflation expectations from Consensus Economics, the estimation sample for this model starts in the '90 and it is evaluated over a shorter period, similarly to the models used in Section 3.1. Figure 11 in the Appendix shows the pseudo real-time estimates of trend inflation from both versions of Chan et al. (2018) model with output gap as well as from Chan et al. (2016) model. We can note that linking the unobserved trend to long-term inflation expectations from surveys results in much less volatile trend estimates, which are also revised less as new observations are added. In the recent period there is sizable difference in the results from the two versions of Chan et al. (2018) model - the trend estimates from *CCK18 no Exp* are much lower.[37]

Figure 8 presents the forecasting performance of the models for the two main slack indicators. This performance is also compared to one of a more traditional model, *Workhorse+EWMA*, that is the workhorse model presented in Section 2 but with a time-varying trend estimated by the EWMA (and estimated in a recursive fashion for comparability). In contrast to the results reported above, only single specifications are evaluated.

We find that, overall, the 'new generation' Phillips curve models can potentially bring forecast gains and are a useful element in a forecaster's toolkit. *CCK18* is better or at least as good as *Workhorse+EWMA* (and much better than the benchmark) for most of the evaluation sample. Its relative performance deteriorates in the most recent period, which we interpret as being related to the linking of the inflation trend to measures of inflation expectations (also in light of the results presented in Figure 4). In that period the version without the link to the expectations (*CCK18 no Exp*) performs better. On the other hand, this version does not improve on the *Workhorse+EWMA*, suggesting that there are rather limited gains from modelling the inflation trend within the model and allowing for

[35] As observed for the recent recovery period in the euro area or the low inflation period in Japan.

[36] We take a burn-in sample of 5000 draws and thereafter we generate 25000 draws from the posterior distribution.

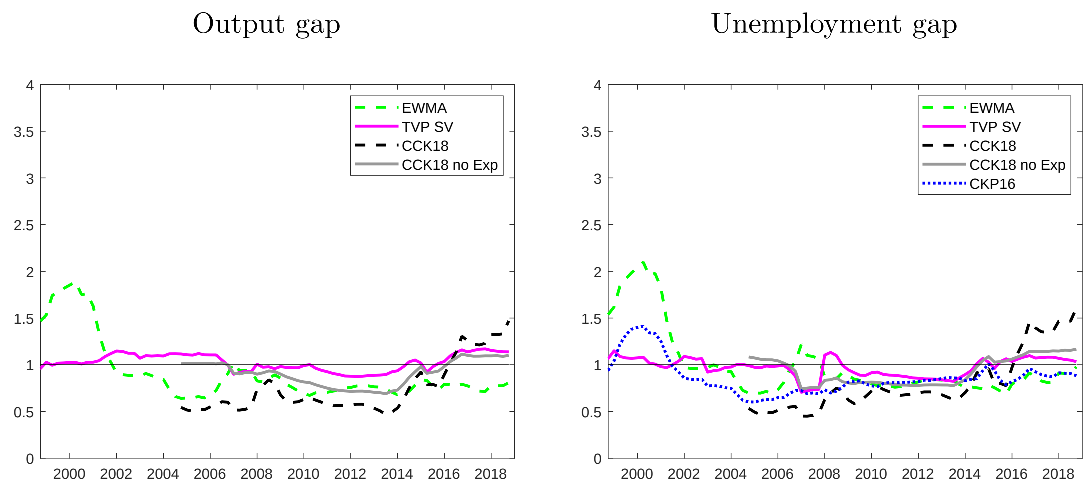
[37] Figure 24 in the Appendix reports the parameter estimates from both versions of Chan et al. (2018) model with output gap based on the full sample. The slope coefficients in both versions are similar. The autoregressive coefficients and stochastic volatility in *CCK18 no Exp* are somewhat lower, reflecting the more volatile trend. The estimates of b for *CCK18* are close to 1 and of a around 0.3, which means that the equation linking the trend to the long-term surveys essentially performs "bias correction" (as in the model with survey expectations detrending in Section 3.1).

time-varying coefficients and stochastic volatility, at least what regards the point forecasts.[38] *CKP16*, with an endogenously estimated unemployment gap, also performs well and, in particular, better than the UCSV towards the end of sample (as opposed to the other models). Thus, it appears that estimating the amount of slack within a Phillips curve model helps to alleviate the challenges related to the assessment of slack (in the labour market) discussed above (see also Jarociński and Lenza, 2018). The *TVP SV* specification performs the worst overall.

What regards the accuracy of density forecast, as indicated by the CRPS and the LPS in Figure 25 in Appendix C, the observations for 'new generation' models remain overall the same but the relative performance of the *Workhorse+EWMA* model deteriorates, especially in terms of the LPS. Also, the probability integral transforms, reported in Figure 26, indicate somewhat worse calibration of this model. This suggests advantages of the 'new generation' models when predictive distributions are of interest.

The forecasts produced by these models are plausible and by and large inflation outcomes lie within the credible intervals, as shown in Figure 9. One can notice that *CCK18* overestimates inflation after 2013 (in line with the behaviour of survey expectations). This is also the case for *Workhorse+EWMA* with unemployment gap albeit to a lesser extent. By contrast the forecasts from *TVP SV*, *CCK18 no Exp* and *CKP16* are largely unbiased, confirming the the better performance of these models over the recent period.

Figure 8: Forecasting performance of new generation Phillips curve models



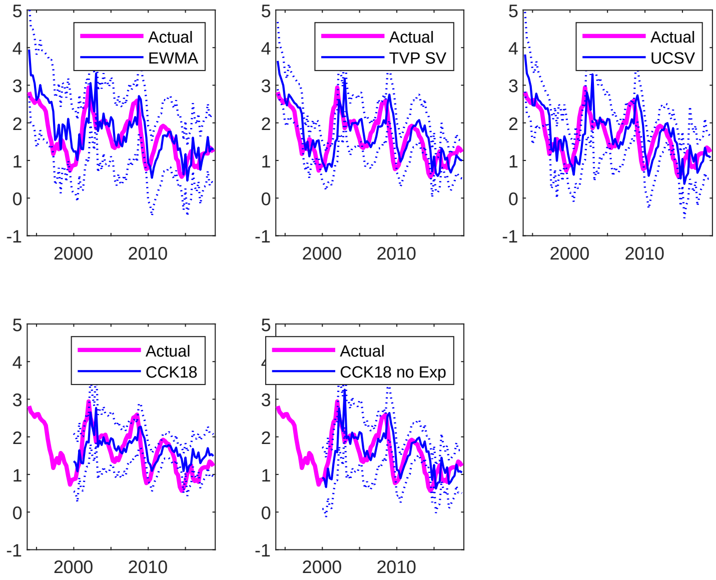
Note: The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. All models are estimated via Bayesian methods and RMSFEs are based on the median inflation forecast.

The new generation Phillips curve models appear to be a promising avenue for conducting forecast.

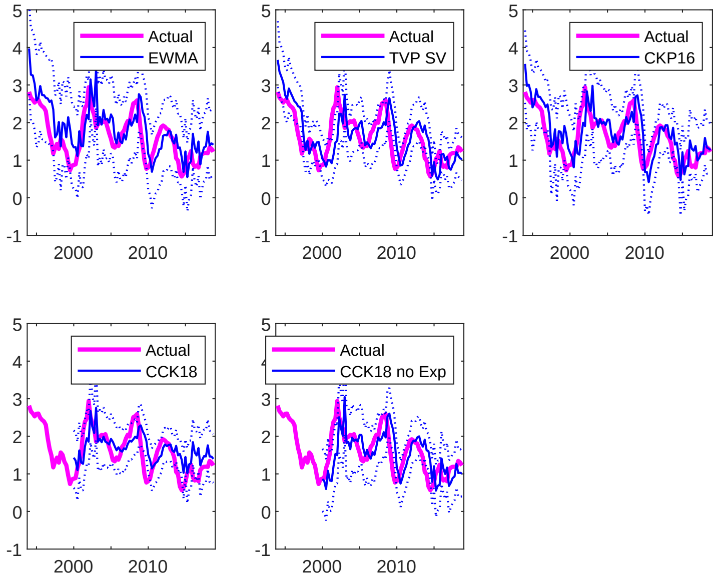
[38]We have also compared *CCK18* and *CCK18 no Exp* to versions where we do not allow the coefficients to vary over time. The RMSFEs remain essentially the same. Also the performance of *CCK18* is similar to the model in which we simply detrend by inflation expectations (see Section 3.1). The results are available upon request.

Figure 9: Actual versus forecasts based on new generation Phillips curve models

Output gap



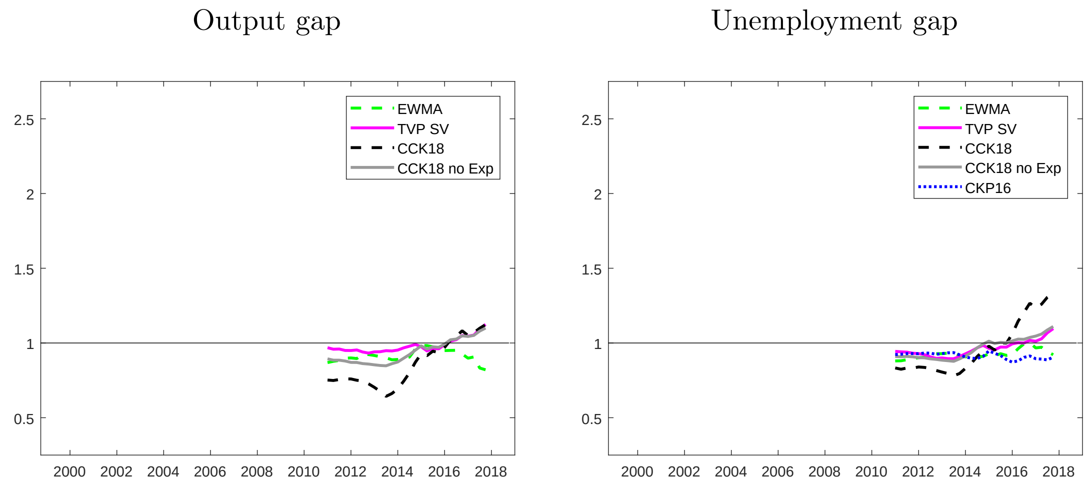
Unemployment gap



Note: The dotted lines show the 80% credible interval.

In order to a have a more complete assessment we also perform a real-time evaluation using the database detailed in Section 3.2.2. [39] Figure 10 shows the results. The lessons drawn using pseudo-real time data are valid also in a real-time set-up. While most models worsen during the euro area post sovereign debt crisis period, the Workhorse + EWMA and CKP16 still outperform the univariate benchmark.

Figure 10: Forecasting performance of new generation Phillips curve models in real time



Note: The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. All models are estimated via Bayesian methods and RMSFEs are based on the median inflation forecast. We use the vintages with the cut-off dates of the Eurosystem/ECB projections.

4 Conclusion

In this paper we evaluate the forecasting performance of a wide range of Phillips curve specifications for euro area HICP inflation excluding energy over the period 1994-2018. Our findings are particularly useful for those building model toolboxes for forecasting inflation and cross-checking nominal and real developments in the euro area economy.

We confirm the well established result (mainly for the US) that when it comes to forecasting inflation, univariate models are hard to beat. In particular, the UCSV model of Stock and Watson (2007) is a tough benchmark and is marginally superior to the random walk of Atkeson and Ohanian (2001). While this is true, some Phillips curve specifications can offer improvements, even if modest, over a univariate benchmark most of the time and are thus a useful element in a forecaster's toolkit.

In other words, measures of economic activity do help to forecast inflation. Unfortunately, this is not a trivial task. As we evaluate models over a long period of time (being the most extensive evaluation for the euro area), we can ascertain that the performance of most simple Phillips curve models is episodic.

Euro area inflation appears to have been harder to forecast when the trend and more recently, also

[39] This evaluation is thus only possible on a shorted sample. We use the vintages with the cut-off dates for the Eurosystem/ECB projections in this exercise.

the amount of slack, were harder to pin down, i.e. in the run-up to the EMU and after the sovereign debt crisis. In both periods, including a time-varying inflation trend in the Phillips curve helps. Actually this is the key type of time-variation to consider; by contrast, the gains from introducing time variation in coefficients are, if at all, modest. In terms of the trend choice, a statistical measure such as an exponential moving average (EWMA) with a low 'forgetting' factor proves to be a good modelling choice, as its gains with respect to the *no detrending* version appear to be more systematic. Long-term inflation expectations from Consensus Economics also appear to be a good gauge of inflation trend in the euro area for most of the sample, with the exception of the 'low inflation period' after the euro area sovereign debt crisis.

The choice of the slack measure also makes a difference. Overall, a simple filter-based measure of the output gap works well, with the exception of the recent years after the double-dip euro area recession. The worsening in the forecast performance is even more visible in the case of a filter-based unemployment gap and we link this to the deep structural changes that affected the euro area economy and especially the labour market in this period , which rendered the estimation of slack more difficult. For this particular period we find that what brings forecast gains is the slack measures produced by international economic institutions, which make use of a large set of economic information to produce their estimates.

'New generation' Phillips curve models incorporating different time-varying features, such as an endogenously estimated trend or slack are a promising model class. They improve on a univariate benchmark most of the time and they are particularly useful when predictive distributions are of interest.

Overall, pooling results from different models and averaging over certain modelling choices and included variables offers some hedge against the instability in the forecast performance.

Important questions for future research arise from the finding that the best performing models after the double-dip recession are typically those with a low inflation trend. This is in line with evidence on declines in trend inflation in the euro area (see e.g. Ciccarelli and Osbat, 2017; Hindrayanto et al., 2019). The analysis does not allow to draw conclusions on the potential drivers of such declines. They could be related to the credibility of the monetary authority, demographics, financial factors, the rise of globalisation or of e-commerce (see e.g. Ciccarelli and Osbat, 2017; Cavallo, 2018). Incorporating such structural drivers into forecasting models has not been thoroughly explored yet.

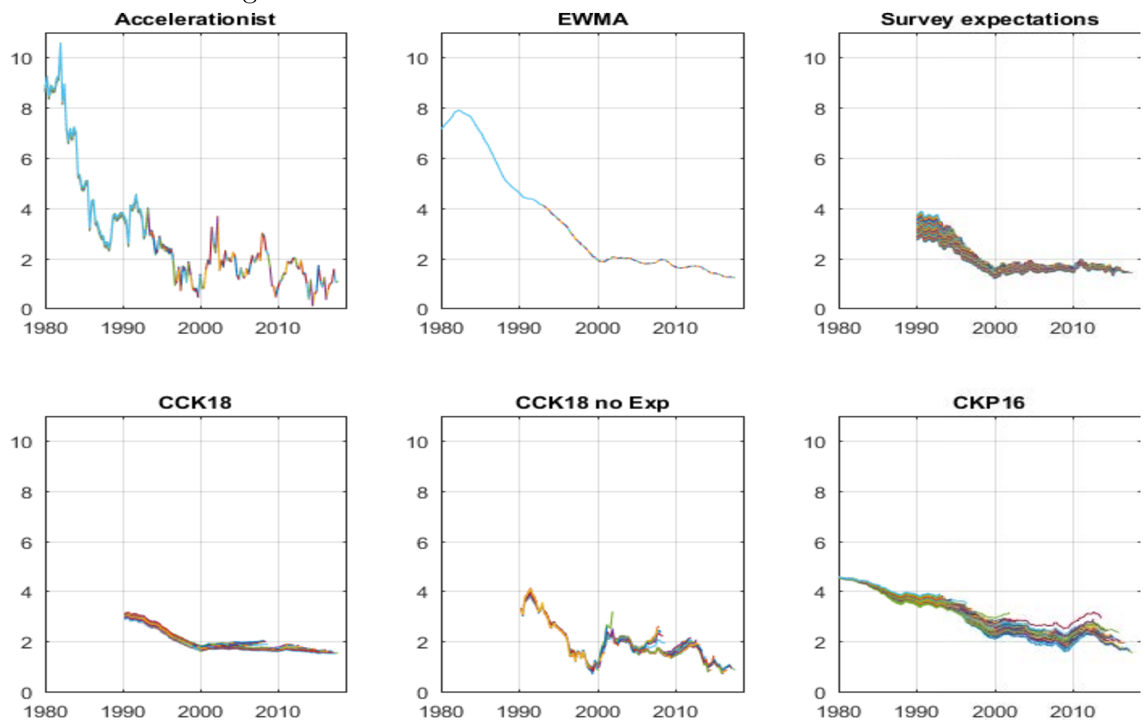
Appendix A Data sources

Variable	Source	Description
		Price index
HICP excluding energy	Eurostat, ECB	Harmonized index of consumer prices, seasonally adjusted
		Inflation expectations
Long-run Consensus expectations	Consensus Economics	6-10 years ahead forecasts for euro area inflation, backcasted
		Domestic real activity variables
Real GDP	Eurostat, AWM database	chain linked volume, calendar and seasonally adjusted, backcasted
Unemployment rate	Eurostat, AWM database	standardised unemployment rate, seasonally adjusted, backcasted
Total employment	Eurostat, AWM database	persons, calendar and seasonally adjusted data, backcasted
Industrial production	ECB	Industrial production index for total industry excluding construction; working day and seasonally adjusted
Capacity utilization	European Commission	survey indicator for total industry, seasonally adjusted, balance of responses
Economic sentiment indicator (ESI)	European Commission	survey indicator, seasonally adjusted, balance of responses
Short term unemployment rate	Eurostat, ECB Staff calculations	percentages, seasonally and working day adjusted; age group 15-74; the short term unemployment rate is the ratio of those who have been unemployed for less than 12 months to labour force. Backcasted by ECB Staff
Broad unemployment rate (U6)	Eurostat, ECB Staff calculations	percentages, seasonally and working day adjusted; age group 15-74; the broad unemployment rate covers, apart from the unemployed, also the underemployed part-time workers, those who are seeking work but are not available and those who are available but are not seeking work (this latter group includes discouraged workers). The sum of these categories is divided by the extended labour force (i.e. the active labour force plus those available, but not seeking work and those seeking work, but not available). Underemployed are part-time workers who would like to work higher hours. Data have been corrected for methodological changes and backcasted by ECB Staff
Principal component	authors' calculatios	first principal component of all the cycles extracted based on the variables above
Output gap augmented with financial variables	based on Melolinna and Toth (2018)	based on an unobserved components model including a financial conditions index, covering: the real growth rate of credit to households, real growth rate of credit to NFCs, real growth rate of M3, real growth rate of residential property prices, spread between short and long term risk free interest rate
		External and externally driven variables
Import deflator	Eurostat, AWM database	deflator of imports of goods and services from the rest of the world, calendar and seasonally adjusted data, backcasted
Import deflator from outside the euro area	Eurostat, AWM database	deflator of imports of goods and services from outside the euro area, calendar and seasonally adjusted data, back-casted
PPI	Eurostat	producer price index for domestic sales, total industry (excluding construction)
US consumer prices	BIS	consumer price index, seasonally adjusted
Price of oil in euro	Bloomberg, ECB	Brent crude oil price converted to euro
Price of non-energy commodities	OECD, ECB	prices of raw materials, excluding energy, converted in euro
Nominal effective exchange rate	ECB	nominal effective exchage rate vis--à--vis 19 trading partners
USD/EUR exchange rate	ECB	exchange rate agaist euro, spot (mid)

Note: Some of the series were backdated using the latest version of the Area-Wide Model (AWM) database, see Fagan et al. (2005). As Consensus expectations are only published at semi-annual frequency we assume that they remain unchanged in the intermediate quarters. They are available for the euro area only as of 2003 and were back-casted to 1990 using the expectations for the largest euro area countries (see Castelnuovo et al., 2003). The short-term unemployment rate and the broad unemployment rate have been back-casted based on a dynamic factor model comprising around 50 variables relevant for labour market dynamics for the whole economy and sectors, such as employment series, hours worked, labour productivity, labour force, unemployment rates (by duration and by cohort), type of employment (self-employed, employees, under-employed), survey variables (ESI, PMI on employment and productivity).

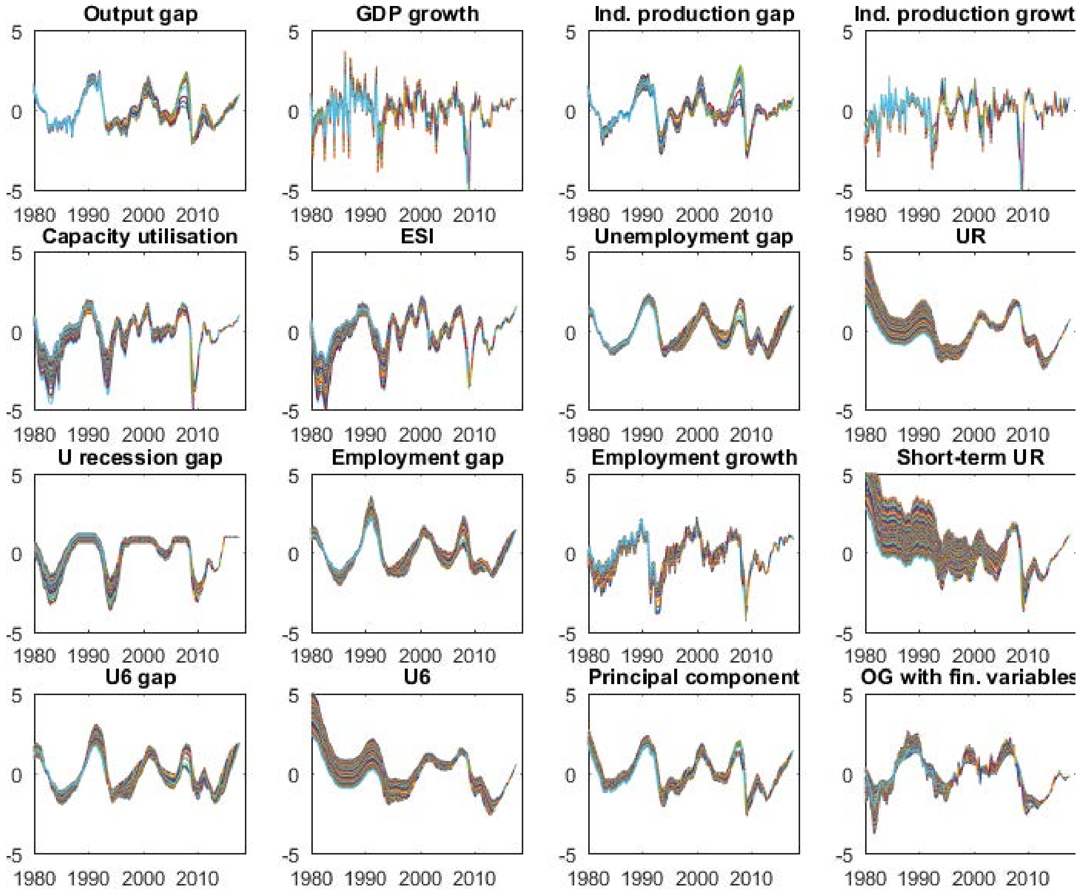
Appendix B Inflation trend and economic slack estimates

Figure 11: Pseudo real-time estimates of inflation trend



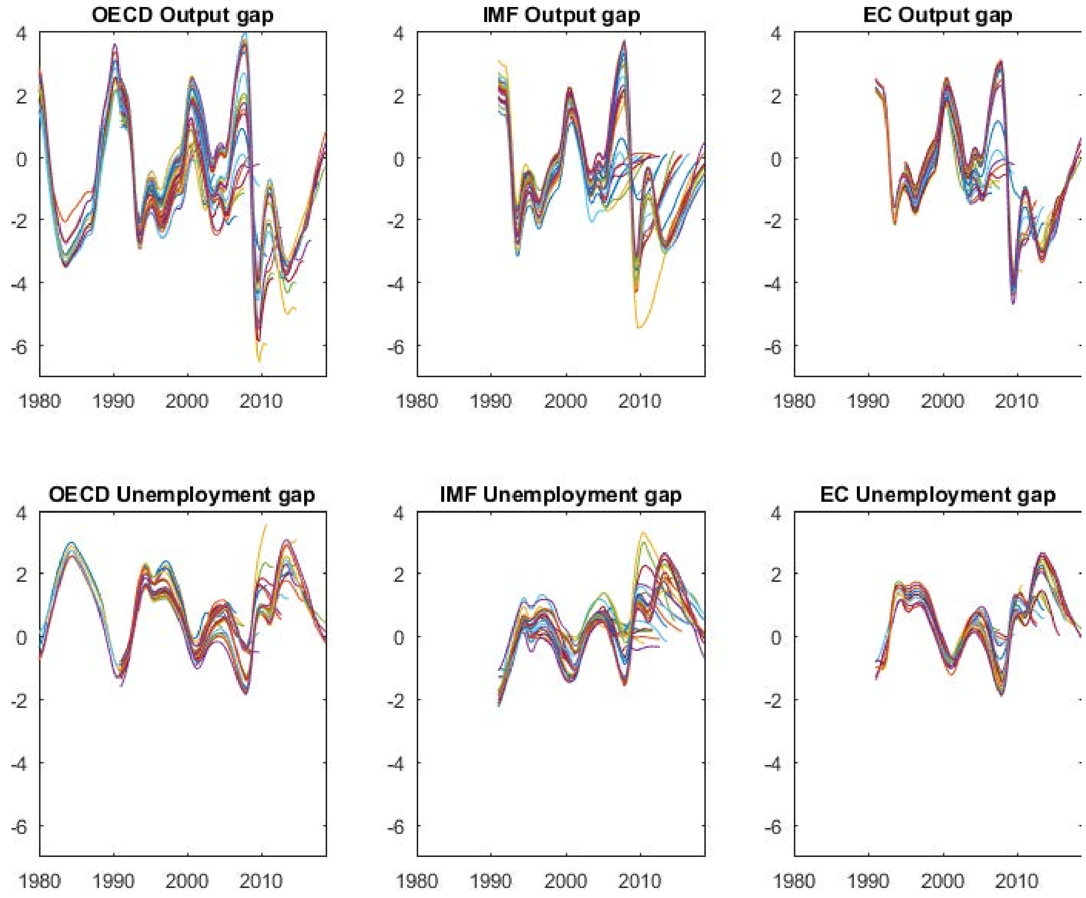
Note: Under the "accelerationist" assumption trend inflation is the past inflation rate. EWMA stands for the exponential weighted moving average inflation trend (with a "smoothing" parameter equal to 0.05). The revisions in the long-term survey expectations are due to the fact that they are corrected for the difference in historical averages between the headline inflation rate (their target variable) and the HICP inflation excluding energy (the target variable in here). *CCK18*, *CCK18 no Exp* and *CKP16* correspond to the recursive estimates of inflation trend obtained from the models described in Section 3.4. The former two correspond to the version with output gap as the measure of slack.

Figure 12: Pseudo real-time estimates of slack



Note: The gaps are obtained by using the Christiano-Fitzgerald filter, keeping the cycles shorter than 15 years. All slack measures are normalised by their standard deviation and those based on the unemployment rate measures are inverted. Slack measures that are not conceptually mean 0 (as is the case for gaps) are demeaned. The means and standard deviations are derived in a "rolling" fashion over a 60 quarter window (corresponding to the setup for the workhorse model). Data revisions are not accounted for. For obtaining the different vintages of the output gap enhanced with financial variables the unobserved components model was estimated in a pseudo real-time fashion.

Figure 13: Real-time estimates of the output and unemployment gap from international institutions

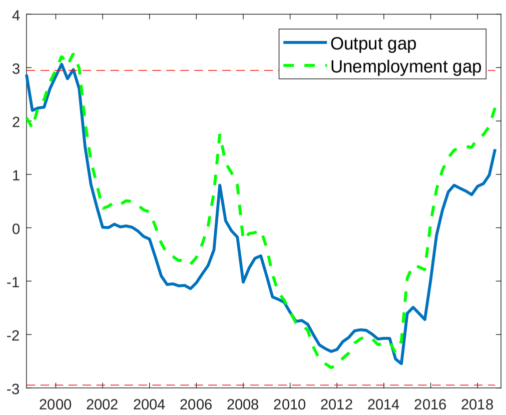


Note: The annual data is interpolated using the Stram-Wei method.

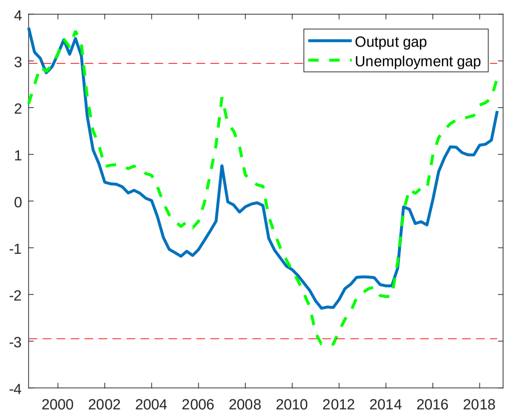
Appendix C Additional results

Figure 14: The Giacomini Rossi fluctuation test for the workhorse model

Benchmark: RW



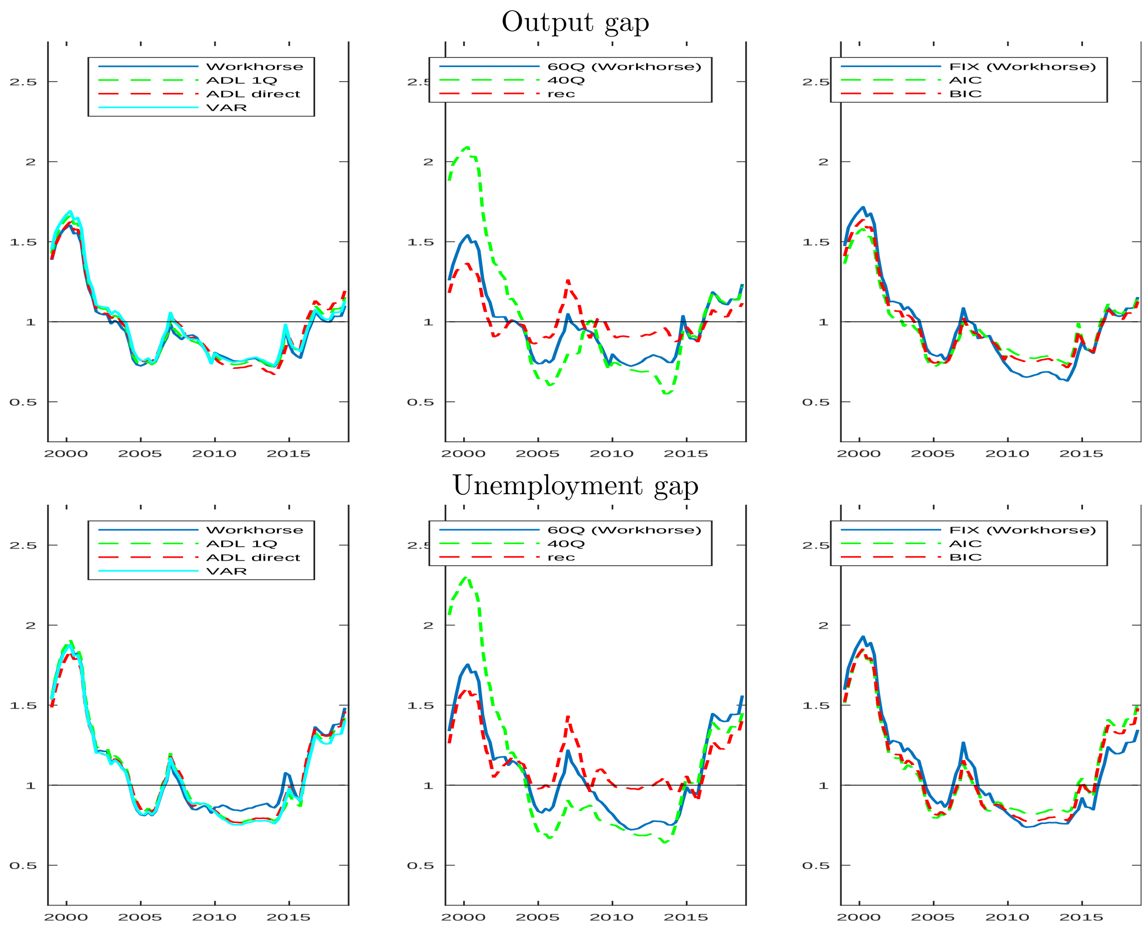
Benchmark: UCSV



Note: Giacomini and Rossi (2010) fluctuation test for a rolling window of 20 quarters. Dashed lines show the critical values for the 90 % credible interval. The null of equal forecasting performance is rejected when the test statistic is outside the critical values interval. Values of test statistic above (below) the interval indicate that the benchmark model is significantly more (less) accurate.

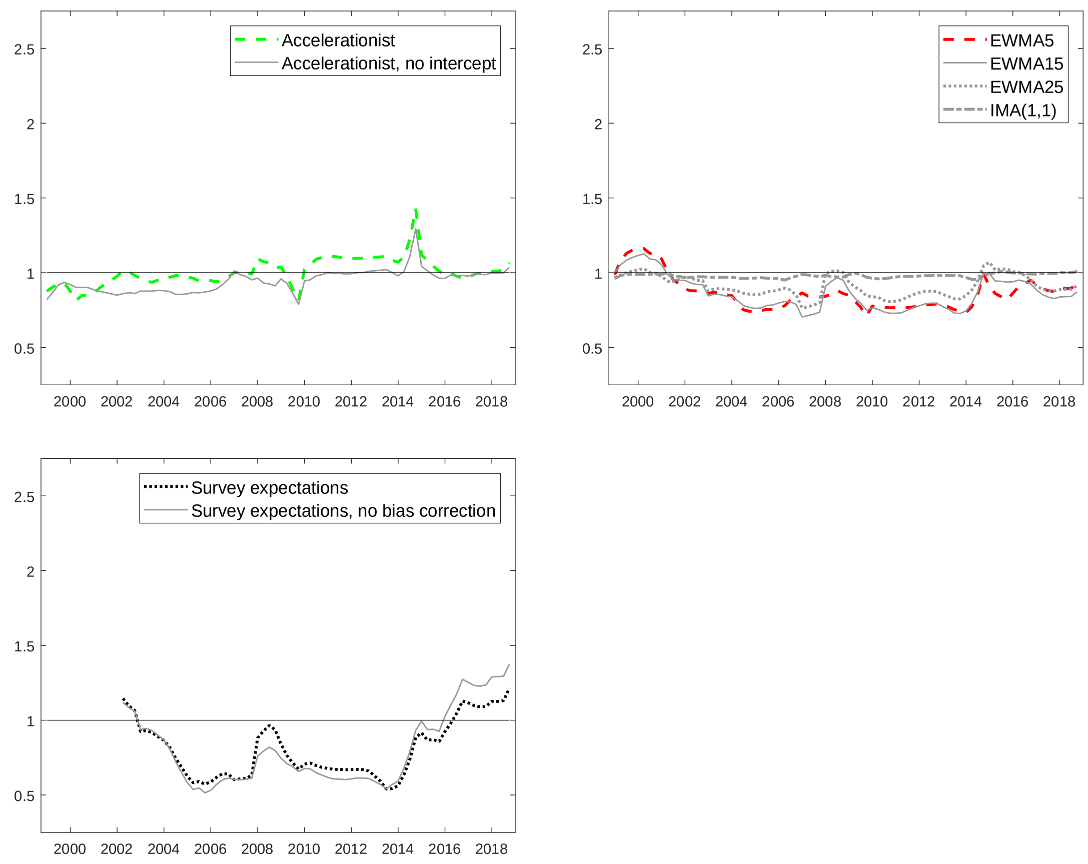
Figure 15: The impact of certain modelling choices on the forecasting performance

Functional form | Estimation sample | Lag selection



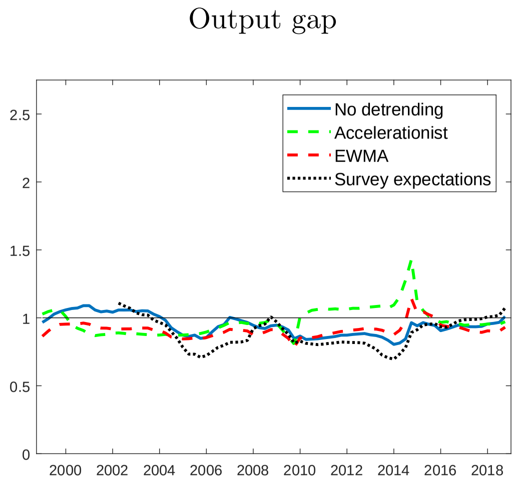
Note: The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. The chart compares accuracy across functional forms (first column), estimation samples (second column) and lag selection approaches (third column). The first column presents accuracy of forecast averages across estimation samples and lag choice methods. The second column refers to averages across functional forms and lag choice methods. The third column refers to averages across functional forms and estimation samples. The difference between the *Workhorse, ADL 1Q* and *ADL direct* resides in the lag structure allowed for slack measures: for the *Workhorse* model this starts with the current quarter, for the *ADL 1Q* it starts with the previous quarter and for *ADL direct* it starts with 4 quarters ago; in all cases it is imposed that at least one lag is kept. In contrast to *ADL direct*, the forecasts for *Workhorse, ADL 1Q* and *VAR* are obtained iteratively. *rec* refers to recursive estimation (starting in 1980).

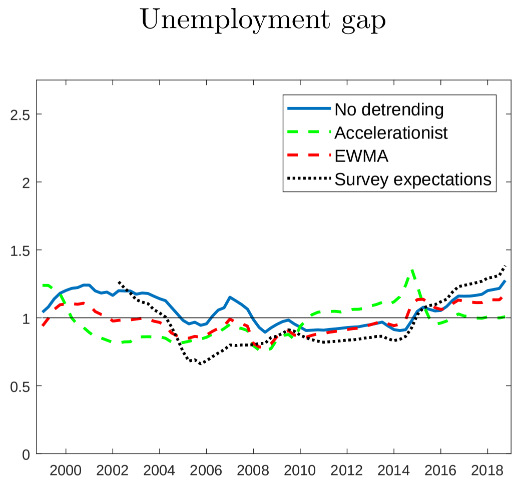
Figure 16: The impact of inflation detrending on the forecasting performance



Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows and lag selection approaches for specifications with the output gap. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. Top left panel: specifications with inflation in differences, with and without intercept, as in Stock and Watson (1999, 2009) and Gordon (1982, 1990), respectively. Top right panel: specifications with EWMA trend with ϕ equal to 0.05, 0.15, 0.25 or estimated from an IMA(1,1) model. Bottom left panel: specifications with trend equal to survey expectations with and without bias correction.

Figure 17: Forecasting performance of the Phillips curve models relative to versions without a slack measure

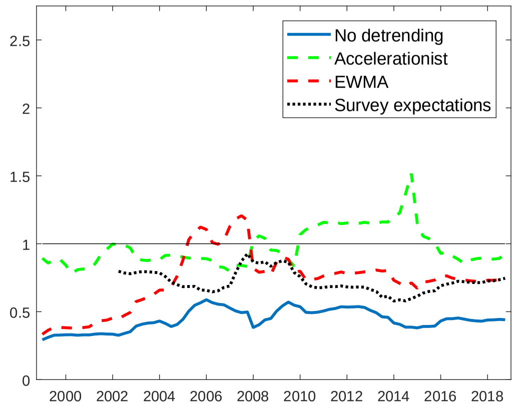




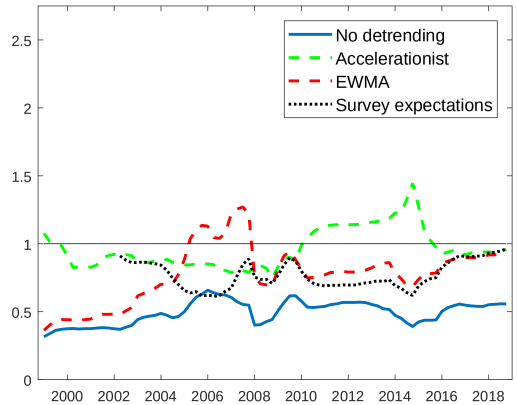
Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of a univariate benchmark, i.e. a version of Equation 2 in which the slope coefficient β is restricted at 0.

Figure 18: Forecasting performance of the Phillips curve models relative to a pure trend forecast

Output gap



Unemployment gap

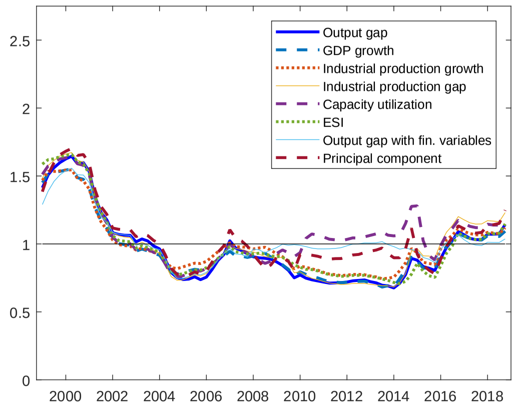


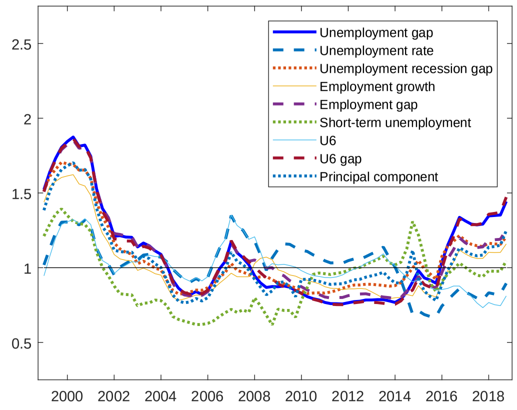
Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of a model where the trend is the forecast, i.e. a version of Equation 2 in which the autoregressive coefficient α and the slope coefficient β are restricted at 0.

Figure 19: The impact of the choice of slack measure on the forecasting performance

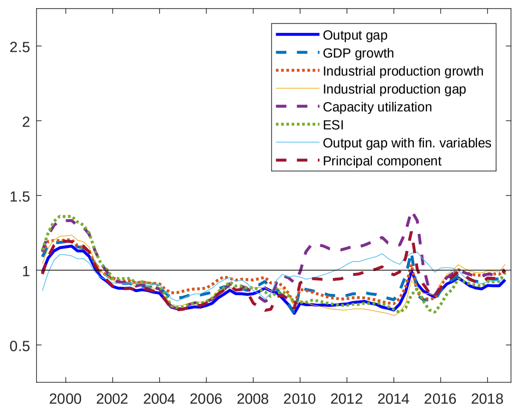
Slack in product market | Slack in labour market

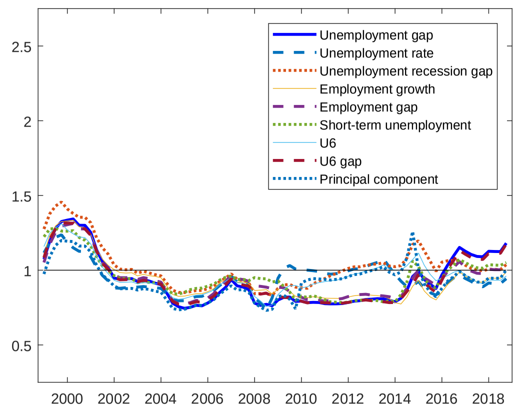
No detrending





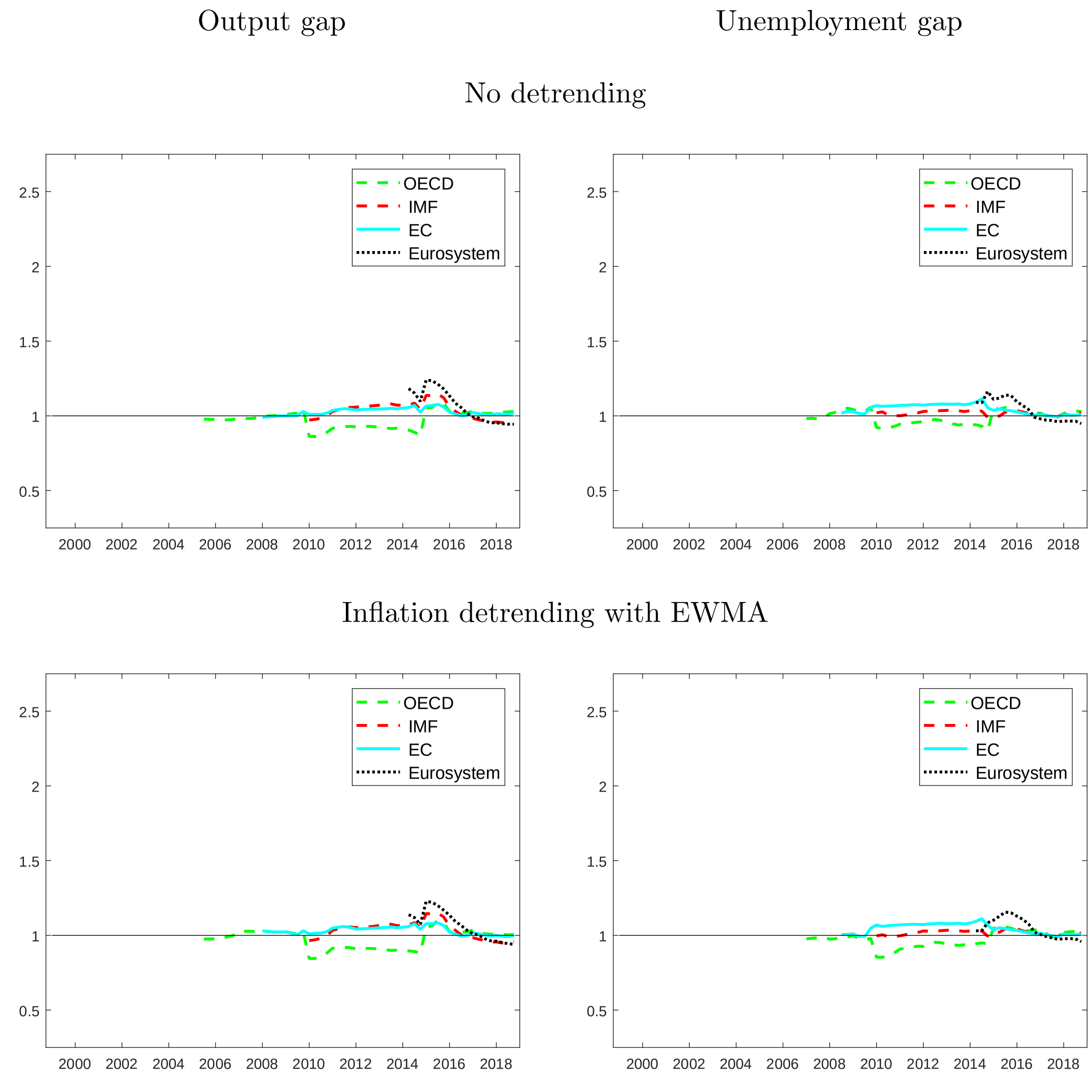
Inflation detrending with EWMA





Note: RMSFEs of average forecast over all considered functional forms, estimation samples and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model.

Figure 20: Accuracy of forecasts *conditional* on projections for slack from international economic institutions relative to their unconditional counterpart

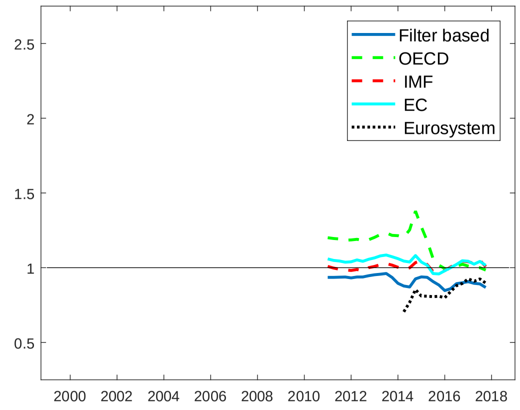


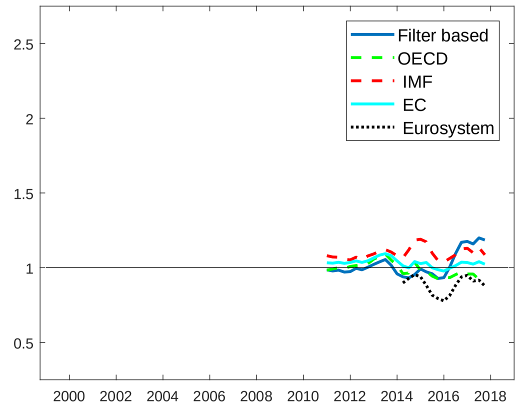
Note: RMSFEs of average forecast over ADL 1Q, VAR and Workhorse model specifications, but also across estimation samples and lag choice approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE from the unconditional forecasts of the same models.

Figure 21: Real-time evaluation of “institutional” slack measures

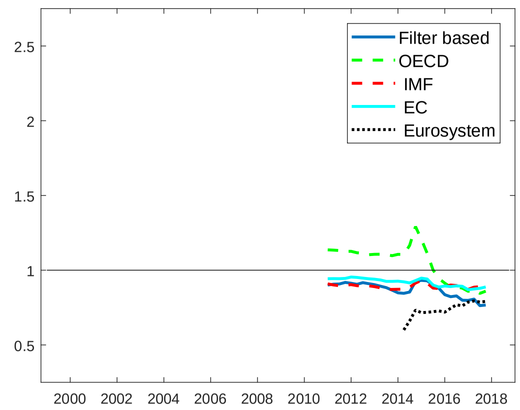
Slack in product market | Slack in labour market

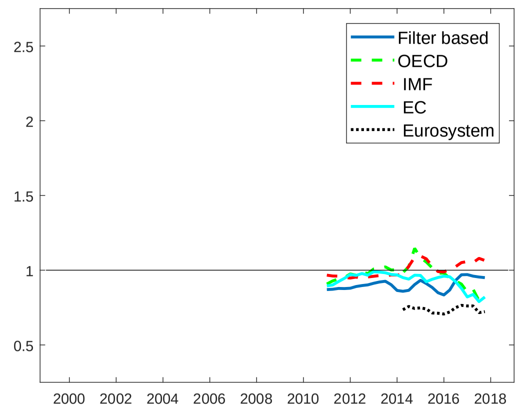
No detrending





Inflation detrending with EWMA





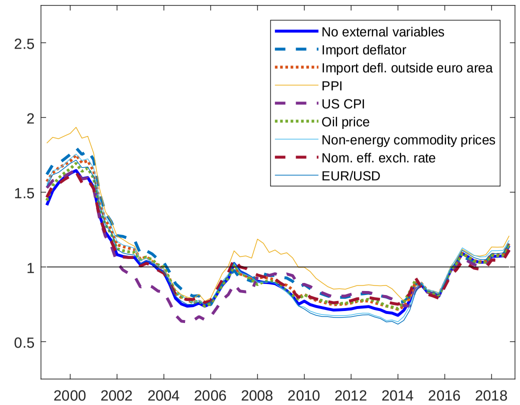
Note: RMSFEs of average forecast over all considered functional forms, estimation samples and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. For filter-based measures of slack we use the vintages with the cut-off dates of the Eurosystem/ECB projections.

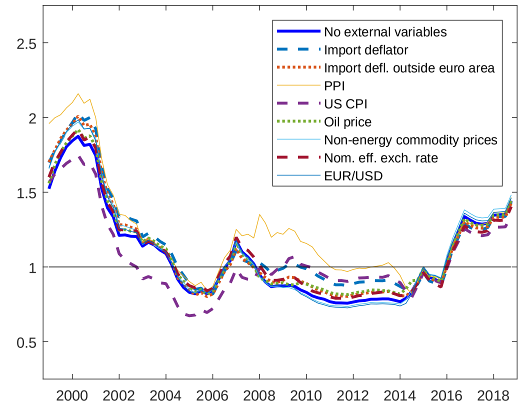
Figure 22: The impact of considering external developments on the forecasting performance

Output gap

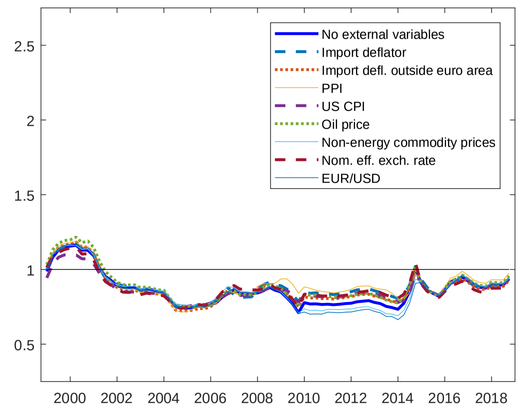
Unemployment gap

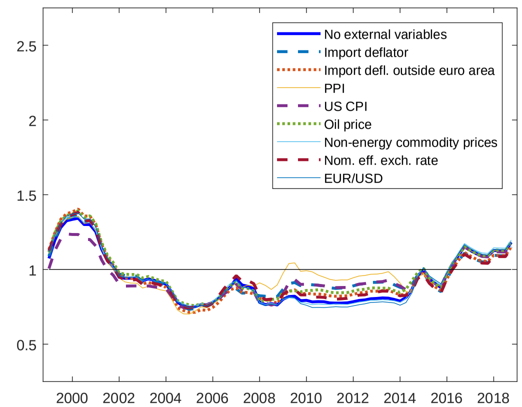
No detrending





Inflation detrending with EWMA

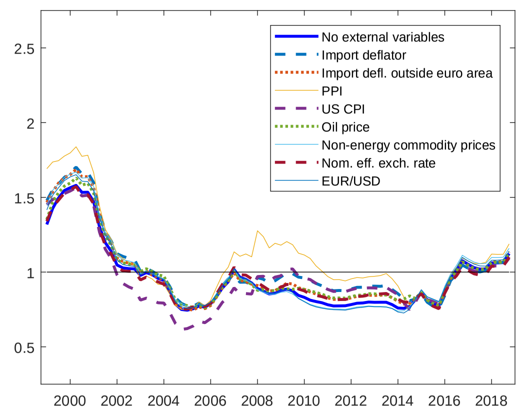




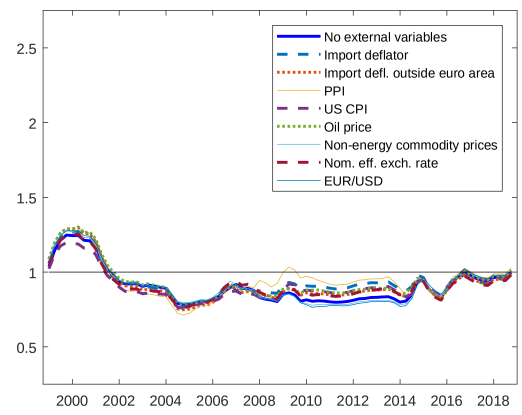
Note: Average RMSFEs over all considered functional forms, estimation windows and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model.

Figure 23: The impact of considering external developments on the forecasting performance

No inflation detrending

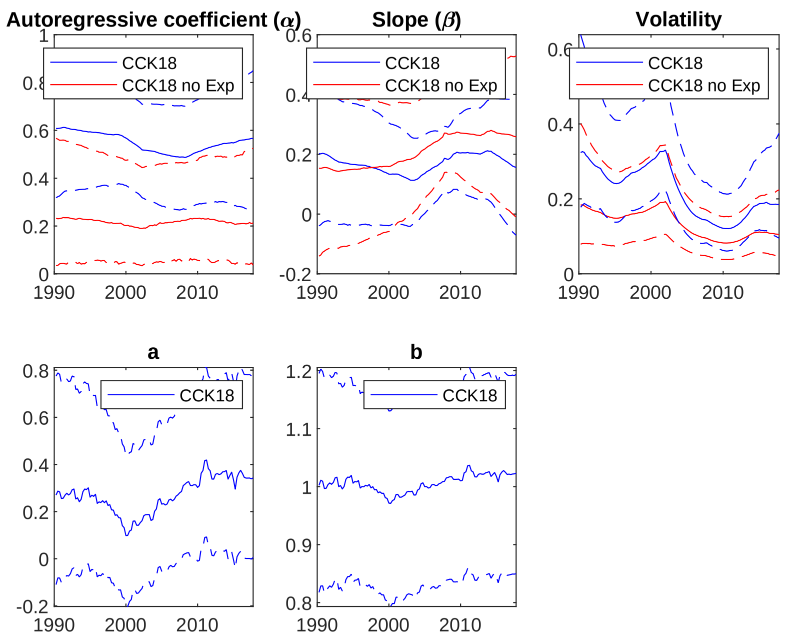


Inflation detrending with EWMA



Note: RMSFEs of average forecast over all considered functional forms, estimation samples, lag selection approaches and slack measures. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model.

Figure 24: Parameter estimates in *CCK18* and *CCK18 no Exp* models.

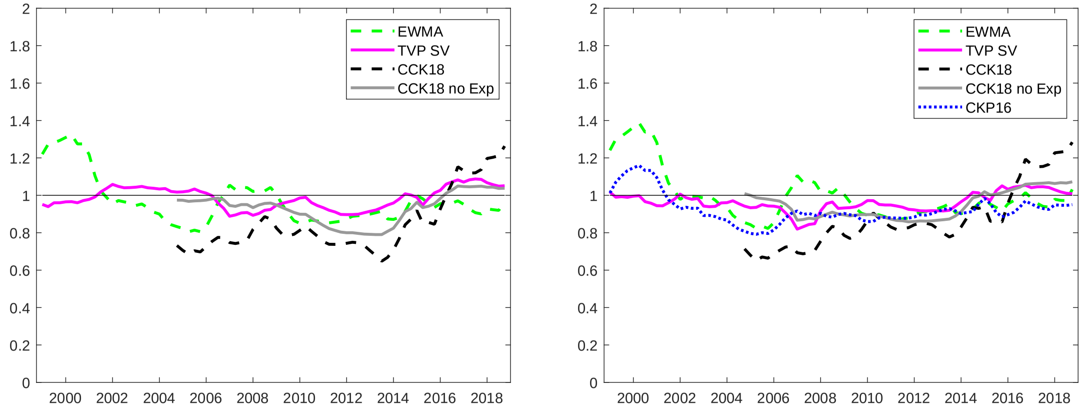


Note: The estimates are for the specification with the output gap and are based on full sample. Solid lines represent the medians and dashed lines the [5% 95%] credible intervals. See Section 3.4 for the description of the models and the notation.

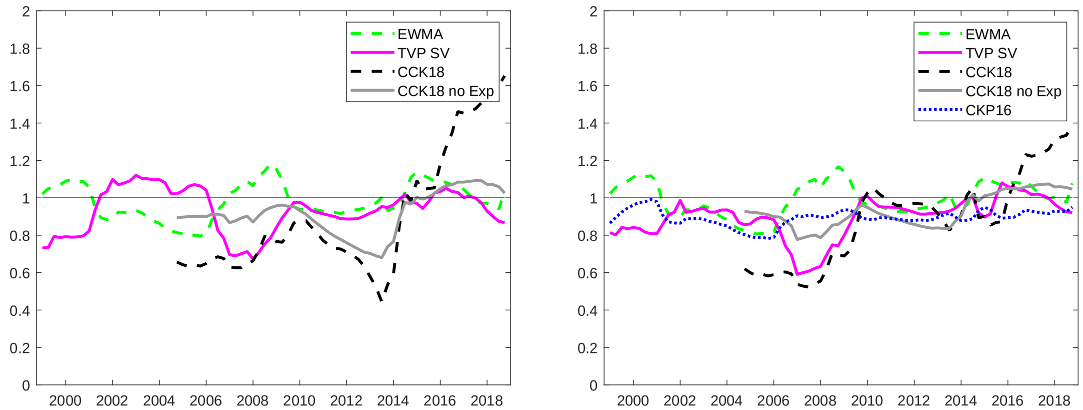
Figure 25: Forecasting performance of the new generation Phillips curve models

Continuous rank probability score

Output gap | Unemployment gap



Log predictive score



Note: The continuous rank probability score and log predictive score respectively were computed over a rolling window of 20 quarters relative to the ones corresponding to the UCSV model.

Figure 26: Forecasting performance of the new generation Phillips curve models.

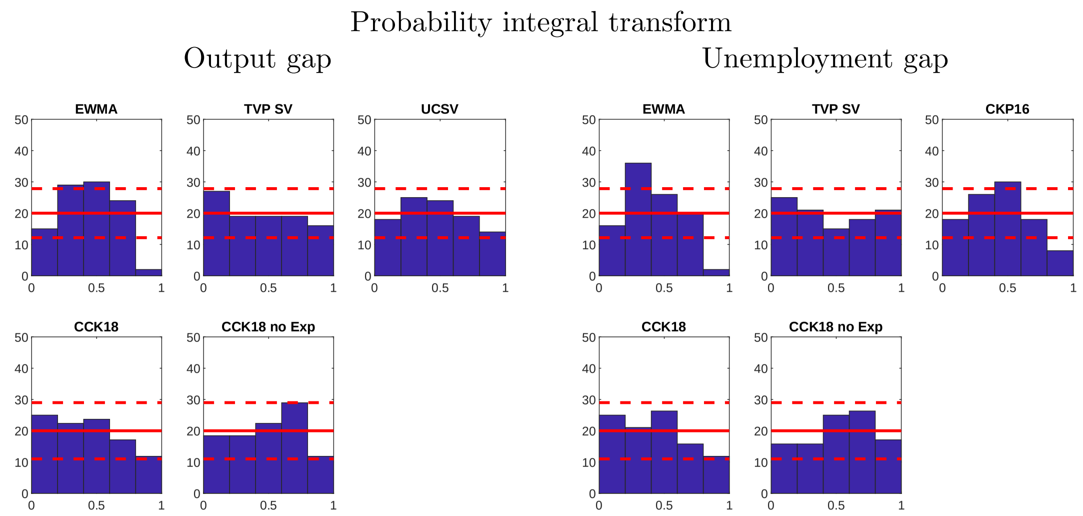
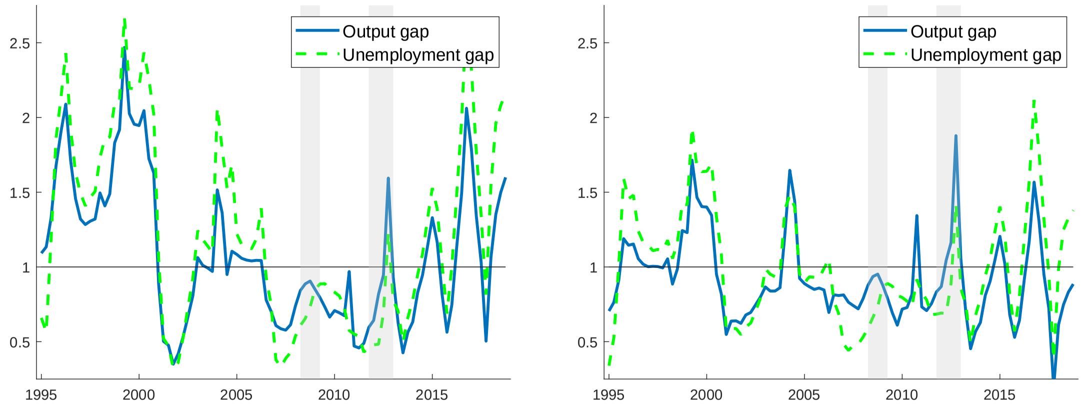


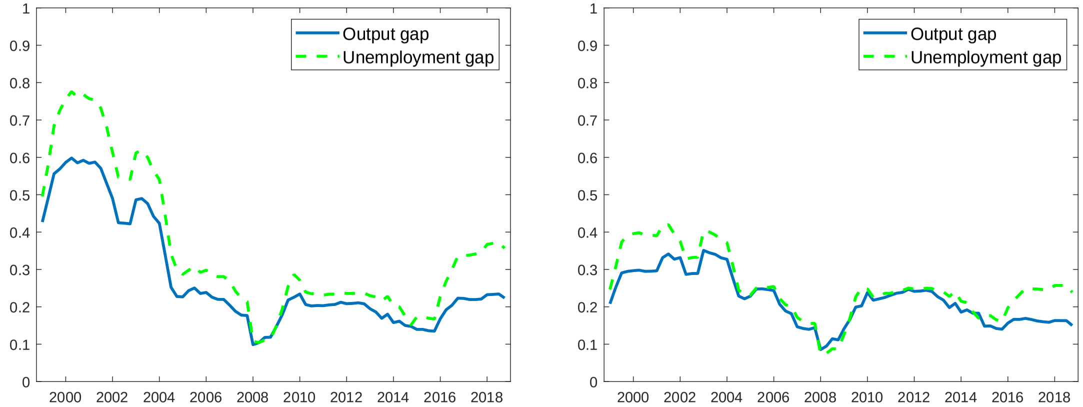
Figure 27: Forecasting performance, additional results

No detrending | Inflation detrending with EWMA

Forecasting performance and recessions



Absolute forecasting performance



Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows and lag selection approaches. In the upper panel the RMSFEs were computed over a rolling window of 4 quarters relative to the corresponding RMSFE of the UCSV model. Shaded areas indicate recession periods as identified by the CEPR Business Cycle Dating Committee. The lower panel shows the (absolute) RMSFEs computed over a rolling window of 20 quarters.

References

Andersson, M., Szörfi, B., Tóth, M., and Zorell, N. (2018). Potential output in the post-crisis period. Economic Bulletin Article 7/2018, European Central Bank.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation. *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1):2–11.

Ball, L. and Mazumder, S. (2011). Inflation dynamics and the Great Recession. *Brookings Papers on Economic Activity*, 42(1):337–405.

Ball, L. and Mazumder, S. (2019a). A Phillips curve with anchored expectations and short-term unemployment. *Journal of Money, Credit and Banking*, 51(1):111–137.

Ball, L. M. and Mazumder, S. (2019b). The nonpuzzling behavior of median inflation. Working Paper 25512, National Bureau of Economic Research.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In Graham Elliott, A. T., editor, *Handbook of Economic Forecasting*, pages 195–237. North Holland.

Bańbura, M. and van Vlodrop, A. (2018). Forecasting with Bayesian vector autoregressions with time variation in the mean. Tinbergen Institute Discussion Papers 18-025/IV, Tinbergen Institute.

Bell, D. N. and Blanchflower, D. G. (2018). Underemployment in the US and Europe. NBER Working Papers 24927, National Bureau of Economic Research, Inc.

Benkovskis, K., Caivano, M., D'Agostino, A., Dieppe, A., Hurtado, S., Karlsson, T., Ortega, E., and Várnai, T. (2011). Assessing the sensitivity of inflation to economic activity. Working Paper Series 1357, European Central Bank.

Bereau, S., Faubert, V., and Schmidt, K. (2018). Explaining and forecasting euro area inflation: the role of domestic and global factors. Working Paper 663, Banque de France.

Blanchard, O. (2016). The US Phillips curve: Back to the 60s? Peterson Institute for International Economics, Working Paper WP15-19.

Blanchard, O., Cerutti, E., and Summers, L. H. (2015). Inflation and activity: Two explorations and their monetary policy implications. Peterson Institute for International Economics, Working Paper WP15-19.

Bobeica, E. and Jarociński, M. (2019). Missing disinflation and missing inflation: A VAR perspective. *International Journal of Central Banking*, 15(1):199–232.

Bobeica, E. and Sokol, A. (2019). Drivers of underlying inflation in the euro area over time: a Phillips curve perspective. Economic Bulletin Article 4/2019, European Central Bank.

Bonam, D., Galati, G., Hindrayanto, I., Hoeberichts, M., Samarina, A., and Stanga, I. (2019). Inflation in the euro area since the global financial crisis. Technical Report 17-3, De Nederlandsche Bank.

Borio, C., Disyatat, P., and Juselius, M. (2015). Rethinking potential output: Embedding information about the financial cycle. PIER Discussion Papers 5., Puey Ungphakorn Institute for Economic Research.

Bracha, A. and Burke, M. A. (2016). Who counts as employed?: informal work, employment status, and labor market slack. Working Papers 16-29, Federal Reserve Bank of Boston.

Buelens, C. (2012). Inflation forecasting and the crisis: Assessing the performance of different forecasting models and methods. *European Comission Economic Papers*, (451).

Canova, F. (2007). G-7 inflation forecasts: Random walk, Phillips curve, or what else? *Macroeconomic Dynamics*, 11(1):1–30.

Castelnuovo, E., Nicoletti-Altimari, S., and Rodríguez-Palenzuela, D. (2003). Definition of price stability, range and point inflation targets: The anchoring of long-term inflation expectations. Working Paper Series 273, European Central Bank.

Cavallo, A. F. (2018). More amazon effects: Online competition and pricing behaviors, Jackson Hole Economic Symposium Conference Proceedings, Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming. Technical report.

Chan, J. C., Clark, T. E., and Koop, G. (2018). A new model of inflation, trend inflation, and long-run inflation expectations. *Journal of Money, Credit and Banking*, 50(1):5–53.

Chan, J. C. C. (2018). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*, 37(8):807–823.

Chan, J. C. C., Koop, G., and Potter, S. M. (2016). A bounded model of time variation in trend inflation, NAIRU and the Phillips curve. *Journal of Applied Econometrics*, 31(3):551–565.

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

Ciccarelli, M. and Osbat, C. (2017). Low inflation in the euro area: Causes and consequences. Occasional Paper Series 181, European Central Bank.

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

Clark, T. E. and McCracken, M. W. (2008). Forecasting with small macroeconomic VARs in the presence of instabilities. volume 3 of *Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization)*, chapter 3, pages 93–147. Emerald Group Publishing Limited, Bingley.

Clark, T. E. and McCracken, M. W. (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 25(1):5–29.

Coeure, B. (2017). Scars or scratches? Hysteresis in the euro area, Speech at the International Center for Monetary and Banking Studies, Geneva, 19 May 2017. 19 may 2017, European Central Bank.

Coibion, O. and Gorodnichenko, Y. (2015). Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics*, 7(1):197–232.

Coibion, O., Gorodnichenko, Y., and Ulate, M. (2019). Is inflation just around the corner? the Phillips curve and global inflationary pressures. NBER Working Papers 25511, National Bureau of Economic Research, Inc.

Constancio, V. (2015). Understanding inflation dynamics and monetary policy. Speech at the Jackson Hole Economic Policy Symposium.

D'Agostino, A., Gambetti, L., and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82–101.

Del Negro, M., Lenza, M., Primiceri, G. E., and Tambalotti, A. (2020). What's up with the Phillips Curve? NBER Working Papers 27003, National Bureau of Economic Research, Inc.

Doepke, J., Dovern, J., Fritsche, U., and Slacalek, J. (2008). Sticky information Phillips curves: European evidence. *Journal of Money, Credit and Banking*, 40(7):1513–1519.

Dotsey, M., Fujita, S., and Stark, T. (2018). Do Phillips curves conditionally help to forecast inflation? *International Journal of Central Banking*, 14(4):43–92.

ECB (2016). A guide to the Eurosystem/ECB staff macroeconomic projection exercises. Technical report, European Central Bank.

Eser, F., Karadi, P., Lane, P. R., Moretti, L., and Osbat, C. (2020). The Phillips Curve at the ECB. Working Paper Series 2400, European Central Bank.

Fagan, G., Hendry, J., and Mestre, R. (2005). An area-wide model for the euro area. *Economic Modelling*, 22(1):39–59.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, pages 2–56. North Holland.

Fisher, J. D. M., Liu, C. T., and Zhou, R. (2002). When can we forecast inflation? *Economic Perspectives - Federal Reserve Bank of Chicago*, 26(1):30–42.

Forbes, K., Kirkham, L., and Theodoridis, K. (2019). A trendy approach to UK inflation dynamics. *The Manchester School*. Forthcoming.

Forbes, K. J. (2019). Inflation dynamics: Dead, dormant, or determined abroad? Nber working papers 26496,, NBER Working Papers 26496, National Bureau of Economic Research, Inc.

Friedman, M. (1968). The role of monetary policy. *American Economic Review*, 58:1–17.

Fuhrer, J. C., Olivei, G. P., and Tootell, G. M. B. (2009). Empirical estimates of changing inflation dynamics. Working Papers 09-4, Federal Reserve Bank of Boston.

Galí, J., Gertler, M., and López-Salido, J. D. (2001). European inflation dynamics. *European Economic Review*, 45(7):1237–1270.

Garratt, A., Mitchell, J., and Vahey, S. P. (2014). Measuring output gap nowcast uncertainty. *International Journal of Forecasting*, 30(2):268–279.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

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.

Gordon, R. (2013). The Phillips curve is alive and well: Inflation and the NAIRU during the slow recovery. Working Paper Series 19390, National Bureau of Economic Research.

Gordon, R. J. (1982). Inflation, flexible exchange rates, and the natural rate of unemployment. In Baily, M. N., editor, *Workers, Jobs and Inflation*, pages 89–158. Washington D.C.: The Brookings Institution.

Gordon, R. J. (1990). The Phillips curve now and then. NBER Working Papers 3393, National Bureau of Economic Research, Inc.

Groen, J. J. J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, 31(1):29–44.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *The Review of Economics and Statistics*, 100(5):831–843.

Hasenzagl, T., Pellegrino, F., Reichlin, L., and Ricco, G. (2018). A model of the Fed's view on inflation. CEPR Discussion Papers 12564, C.E.P.R. Discussion Papers.

Hasenzagl, T., Pellegrino, F., Reichlin, L., and Ricco, G. (2019). The inflation puzzle in the euro area – it's the trend not the cycle! Voxeu column, 16 october 2019.

Hindrayanto, I., Samarina, A., and Stanga, I. M. (2019). Is the Phillips curve still alive? evidence from the euro area. *Economics Letters*, 174:149 – 152.

Hooper, P., Mishkin, F. S., and Sufi, A. (2019). Prospects for inflation in a high pressure economy: Is the Phillips curve dead or is it just hibernating? Working Paper 25792, National Bureau of Economic Research.

Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? *International Journal of Forecasting*, 21(1):119–136.

IMF (2013). Chapter 3, The dog that didn't bark: has inflation been muzzled or was it just sleeping? In *World Economic Outlook, April*, pages 1–17. International Monetary Fund, Washington.

Jarociński, M. and Lenza, M. (2018). An inflation-predicting measure of the output gap in the euro area. *Journal of Money, Credit and Banking*, 50(6):1189–1224.

Kozicki, S. and Tinsley, P. A. (2001). Shifting endpoints in the term structure of interest rates. *Journal of monetary economics*, 47(3):613–652.

Lane, P. (2019). The Phillips Curve at the ECB. Speech at the 50th Anniversary Conference of the Money, Macro & Finance Research Group, London School of Economics. 4 september 2019.

Marcellino, M. and Musso, A. (2010). The forecasting performance of real time estimates of the euro area output gap. CEPR Discussion Papers 7763.

McLeay, M. and Tenreyro, S. (2019). Optimal inflation and the identification of the Phillips Curve. In *NBER Macroeconomics Annual 2019, volume 34*, NBER Chapters, pages 199–255. National Bureau of Economic Research, Inc.

Melolinna, M. and Tóth, M. (2018). Output gaps, inflation and financial cycles in the United Kingdom. *Empirical Economics*, 56(4):1039–1070.

Montoya, L. A. and Döhring, B. (2011). The improbable renaissance of the Phillips curve: The crisis and euro area inflation dynamics. European Economy. Economic Papers 446, European Commission.

Moretti, L., Onorante, L., and Zakipour-Saber, S. (2019). Phillips curves in the euro area. Working Paper Series 2295, European Central Bank.

Musso, A., Stracca, L., and van Dijk, D. (2009). Instability and nonlinearity in the euro area Phillips curve. *International Journal of Central Banking*, 5(2):181–212.

O'Reilly, G. and Whelan, K. (2005). Has euro area inflation persistence changed over time. *Review of Economics and Statistics*, 87(4):709–720.

Orphanides, A. and van Norden, S. (2005). The reliability of inflation forecasts based on output gap estimates in real time. *Journal of Money, Credit and Banking*, 37(3):583–601.

Paloviita, M. (2008). Comparing alternative Phillips curve specifications: European results with survey-based expectations. *Applied Economics*, 40(17):2259–2270.

Peneva, E. V. and Rudd, J. B. (2017). The passthrough of labor costs to price inflation. *Journal of Money, Credit and Banking*, 49(8):1777–1802.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134 – 161.

Powell, J. H. (2018). Monetary policy and risk management at a time of low inflation and low unemployment", Speech at the 'Revolution or evolution? reexamining economic paradigms' 60th annual meeting of the national association for business economics, boston, massachusetts. 2 October 2018.

Powell, J. H. (2020). New economic challenges and the fed's monetary policy review, Speech at 'Navigating the decade ahead: Implications for monetary policy,' a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming. 27 august 2018, Federal Reserve System.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.

Riggi, M. and Venditti, F. (2015). Failing to forecast low inflation and phillips curve instability: A euro-area perspective. *International Finance*, 18(1):47–68.

Rossi, B. and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3):432–453.

Rünstler, G. (2002). The information content of real-time output gap estimates: An application to the euro area. Working Paper Series 182, European Central Bank.

Stella, A. and Stock, J. H. (2015). A state-dependent model for inflation forecasting. In Koopman, S. J. and Shephard, N., editors, *Unobserved Components and Time Series Econometrics*. Oxford University Press.

Stevens, A. and Wauters, J. (2018). Is euro area lowflation here to stay ? Insights from a time-varying parameter model with survey data. Working Paper Research 355, National Bank of Belgium.

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

Stock, J. H. and Watson, M. M. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking*, 39(7):3–34.

Stock, J. H. and Watson, M. M. (2009). Phillips curve inflation forecasts. In Fuhrer, J., Kodrzycki, Y., Little, J. S., and Olivei, G., editors, *Understanding Inflation and the Implications for Monetary Policy, a Phillips Curve Retrospective*, pages 101–186. The MIT Press.

Stock, J. H. and Watson, M. W. (2010). Modeling inflation after the crisis. NBER Working Papers 16488, National Bureau of Economic Research, Inc.

Stock, J. H. and Watson, M. W. (2019). Slack and cyclically sensitive inflation. Working Paper 25987, National Bureau of Economic Research.

Wright, J. H. (2013). Evaluating real-time VAR forecasts with an informative democratic prior. *Journal of Applied Econometrics*, (28):762–776.

Yellen, J. (2015). Inflation dynamics and monetary policy, Speech at the Philip Gamble Memorial Lecture, University of Massachusetts, Amherst, Massachusetts. 24 september 2015, Federal Reserve System.

Yellen, J. (2019). Speech at a public event,"What's (not) up with inflation?" hosted Hutchins Center on Fiscal & Monetary Policy at Brookings . 3 october 2019.

Zaman, S. (2013). Improving inflation forecasts in the medium to long term. Economic commentary, Federal Reserve Bank of Cleveland.

Acknowledgements

We thank Laurence Ball, Joshua Chan, Marek Jarociński, Simone Manganelli, Barbara Rossi, Oreste Tristani, Saeed Zaman and participants of the 2nd Forecasting at Central Banks Conference hosted by Bank of England, for valuable comments. We are grateful to Harun Mirza for collaboration at the initial stage of the project; to Katalin Bodnar, Bela Szörfi and Máté Tóth for sharing data on slack measures; and to Joshua Chan and Aubrey Poon for sharing their codes. The views expressed in this paper are of the authors only and do not necessarily reflect those of the European Central Bank or the Eurosystem.

Marta Bańbura
European Central Bank, Frankfurt am Main, Germany; email: marta.banbura@ecb.europa.eu

Elena Bobeica
European Central Bank, Frankfurt am Main, Germany; email: elena.bobeica@ecb.europa.eu

PDF ISBN 978-92-899-4388-8 ISSN 1725-2806 doi:10.2866/61538 QB-AR-20-123-EN-N